DTIC FILE COPY

(4)

AD-A197 250

Report Tantalus/CT-88/3

SENSOR CALCULUS

Considerations in the Design of Distributed Systems for Detection, Discrimination and Decision

Paul B. Kantor
Richard Blankenbecler
Moula Cherikh

Tantalus, Inc.
3257 Ormond Road
Cleveland Heights, OH 44118

17 June 1988

Final Report for 1987-1988
Contract # N00014-87-C-0695

Prepared for
Office of Naval Research
Department of the Navy
800 N. Quincy Street
Arlington, VA 22217

This report is UNCLASSIFIED

Distribution is Unlimited

88

SECURITY CLASSIFICATION OF THIS PAGE This page is Unclassified

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) N/A	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Tantalus; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION ONR Office of Naval Research
6c. ADDRESS (City, State and ZIP Code) 3257 Ormond Rd Cleveland Ohio 44118	7b. ADDRESS (City, State and ZIP Code) 800 N. Quincy Street Arlington VA 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION ONR; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract # N00014-87-C-0695
8c. ADDRESS (City, State and ZIP Code) Department of the Navy 800 N. Quincy Street Arlington, VA 22217	10. SOURCE OF FUNDING NOS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT NO
11. TITLE (Include Security Classification) SENSOR CALCULUS	

12. PERSONAL AUTHOR(S)
Paul B. Kantor, Richard Blankenbecler, Moula Cherikh

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 7/87 TO 6/88	17 June 1988	86

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	detector rule, distributed sensor system, fusion rule, hypothesis, management, (over)
12	04		
25	05		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research applies optimal control and experimental design theories to management of a multidetector system including the cost of false alarms, misses, communications and improvements in operating characteristics. The successful interaction of multiple detectors is applicable to battle management and control.

A powerful new framework is presented for the analysis of distributed detection networks, supported by a compact notation for the description of complicated networks. The new methods are applicable to problems with any number of threats, any number of messages and any number of available actions. As illustrations the Sensor Calculus is applied to reveal interesting features of the case of two-fold threats, messages and actions, including symmetry breaking and suboptimality of — (over)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Paul B. Kantor	22b. TELEPHONE NUMBER (Include Area Code) 216 321 7713; 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

Continuations:

Block 18. Subject terms:
maximum entropy principle, multidetector systems, optimal control theory, probability, programming, signaling and action set, theory of experimental design.

Block 19. Abstract:
deterministic tuning for fusion systems.

SECURITY CLASSIFICATION OF THIS PAGE

Figure 8. Report Documentation Page, DD Form 1473 (2 of 2)

SENSOR CALCULUS

CONSIDERATIONS IN THE DESIGN OF DISTRIBUTED SYSTEMS FOR DETECTION, DISCRIMINATION AND DECISION

Paul B. Kantor, Tantalus, Inc.

Richard Blankenbecler, Consultant

Moula Cherikh, Tantalus, Inc. and Case-Western Reserve University

SUMMARY

A powerful new framework is presented for the analysis of distributed detection networks, supported by a compact notation for the description of complicated networks. The new methods are applicable to problems with any number of threats, any number of messages, and any number of available actions. As illustrations, the Sensor Calculus is applied to reveal some interesting features of the *case of two-fold threats, messages and possible actions.* These features include the occurrence of spontaneous symmetry breaking with identical sensors, and the sub-optimality of deterministic tuning for fusion systems.

FOREWORD

This report covers the first year of a three-year study of the application of optimal control theory to the design of distributed sensor systems. The work is focused on the key links between detection, discrimination and decision. Detection is an engineering/physics problem. Discrimination is affected by softer considerations such as estimates of prior probabilities, which depend on intelligence as well as engineering information. Decision involves even softer estimates of the costs and values associated with various possible damage or loss. Thus the Detection-Discrimination-Decision (D^3) problem spans a range from hard facts to volatile speculation.

Our work concentrates on building a rigorous framework in which hard data serve to define an operating characteristic, and softer data are used to define the optimal tuning of a sensor system. Improvements in detection and discrimination always carry price tags. In the D^3 framework the benefits (in damage control) may be weighed against those costs, for best system management.

The principal results of the first year's work are:

(1) Formalization of the concept of a detector operating characteristic (doc), in a form correct for generalization to any number of possible threats, any number of available responses, and any channel message carrying capacity.

(2) Development of a powerful and compact notation suitable for describing any network of sensors and for determining its doc and, as appropriate, its optimal tuning. (The Sensor Calculus).

Using these tools we have established a number of significant specific results for the simplest case (one threat, one response, and binary message schemes.) Among these results are:

(1) Spontaneous symmetry breaking. It is often found that when two identical sensors are used to inform a fusion (decision) center, their optimal tunings are the same. We have established, by specific examples, that this is not, in general, true. It can be the case that symmetrical tuning is less effective than a suitable symmetry-breaking choice of

tuning. This significantly increases the complexity of finding optimal solutions, but permits improvement in overall performance of the detection network.

(2) Discontinuity of optimal tuning in fusion. We have established that, quite generally, in fusion systems, the optimal tuning may be discontinuous as a function of the softer parameters such as prior probabilities and estimated cost. This has serious implications for optimal system design because the soft parameters are subject to significant change after a system has been constructed. Every effort must be made to ensure that the likely range of variability does not include tuning discontinuities.

(3) When a sensor communicates over a limited channel there is a loss of information. If one sensor is better than another, should the better one send or receive the message over a limited channel? Using the sensor calculus techniques we have established that there is no general rule governing this. In some situations one alternative is better, and in other situations the other is better.

(4) The best achievable deterministic architecture will be significantly suboptimal in some resource-constrained situations. This has serious implications for the allocation of resources among interceptors and sensors.

In all of this analysis the ability to move easily from a discrete to a continuous formulation has enormously clarified our understanding of the problem. We are firmly convinced that reliance on analytical approximations is an artificial and dangerous restriction in the study of D^3 problems.

Our plans for the second and third years of the project are to continue this line of research by: (1) developing algorithms that will accomplish the basic operations of the sensor calculus as efficiently as possible; (2) extending the results to the case of more than two possible states of nature (as, for example, when there may be a variety of decoy threats); (3) extending the results to the case of more than two possible actions and/or messages between sensors; (4) extending the formalism to deal with "call-back" systems in which some message combinations may result in a polling of the sensors; (5) exploration of the implications of the maximum entropy principle for scheduling such "call-backs". The overall goal of the research is to bring the task of combining sensor characteristics to a highly automated state, so that the

consideration of alternative architectures will be reduced to "cook-book" calculations using the calculus of sensors.

Acknowledgements: It is a pleasure to acknowledge stimulating conversations with Dr. Keith Taggart, SDIO/CMO, Jason Goodfriend, Joshua Scharf and R. Barry Thomas at System Planning Corporation, and Dr. Rabindar Madan at the Office of Naval Research.

CONTENTS

1. Introduction and Notation.

This report is part of an ongoing effort to resolve the problem of detection, discrimination and decision (D^3 problem) into design of the Detection-Discrimination network on the one hand, and discussion of the Decision aspects on the other. We find that the natural link between these areas is given by a powerful construct termed the doc (for detector operating characteristic.) The doc generalizes the notion of the ROC (Receiver Operating Characteristic), which is the boundary of the doc in the familiar cases. The presentation is in two main parts: the first deals with the Detection-Discrimination Network; the second deals with the decision problem. We find that the doc plays a central role by (i) describing the overall characteristics of the network for use in the decision problem and by (ii) providing the necessary and sufficient information about each sensor to support solution of the problem of optimal network design. A review of related literature is included as Appendix B.

1.2 Notation.

In Part I we show how a sensor S can be fully characterized by a set of points $\mathfrak{D}(S)$ called the doc of S. We show how the doc is built up from the signal set Y using the response functions $f_1(y)$, ... $f_H(y)$ corresponding to some exclusive and exhaustive list of alternate hypotheses about the world, $h=1,....,H$. We show that the familiar ROC is, in some sense, the boundary of the doc. We introduce the useful concepts of the full product of sensors S and S', $S \otimes S'$ and of the M-fold restriction of a sensor, $\overline{\mathfrak{R}(M)S}$. This latter concept is useful because communication within sensor networks is constrained by the capacities of communication channels.

1.3 Definitions

Although our presentation will not be highly formal, we state here the definitions of the key concepts of the sensor calculus.

A Sensor (S) consists of a signal set Y, which may be discrete or continuous, and a collection of non-negative conditional probability measures defined on Y, $df_1,\ldots,df_H$, corresponding to the possible states of nature.

The detector operating characteristic (doc), $\mathfrak{D}(S)$ is a set of points in an H-dimensional linear vector space, consisting of all points of the form $f_1(\mathbf{Y}(t)),\ldots,f_H(\mathbf{Y}(t))$ where $\mathbf{Y}(t)$ is any measurable subset of $\mathbf{Y}$ and $f_h(\mathbf{Y}(t))$ is the sum or integral of the measure f_h over the set $\mathbf{Y}(t)$. The doc $\mathfrak{D}$ always lies within the closed unit hypercube in the positive orthant determined by the origin and the point $E=(1,\ldots,1)$.

The boundary of the doc of S, $\mathfrak{B}(S)$ is defined as the set of extreme points of $\mathfrak{D}(S)$. A point $P\epsilon\mathfrak{D}$ is an extreme point if there exists a separating hyperplane, $\{x\colon n\cdot x=c\}$ such that $\mathbf{n}\cdot P=c$, and $\mathbf{n}\cdot x\geq c$ for all $x\epsilon\mathfrak{D}(S)$.

An M-fold restriction with tuning **t**, $\mathfrak{R}(M,t)S$ is defined in terms of its doc. The tuning **t** defines a partition of **Y** into M sets $\{\mathbf{Y}(m)\}_{m=1,\ldots,M}$. The doc $\mathfrak{D}(\mathfrak{R}(M,t)S)$ is the discrete set consisting of all the points $f(\mathbf{Y}(m))$. This corresponds to using the sensor S to select one from a set of M options, which may be actions or messages.

<u>1.4 Examples</u>

To illustrate the notation we consider the networks shown in Figure 1. The corresponding expressions in the calculus of sensors are:

Figure 1a. $\mathfrak{R}(2)\{\overline{\mathfrak{R}(2)S}_1\otimes\overline{\mathfrak{R}(2)S}_2\}$

Figure 1b. $\mathfrak{R}(2)\left\{S_3\otimes\mathfrak{R}(2)\{\overline{\mathfrak{R}(2)S}_1\otimes\overline{\mathfrak{R}(2)S}_2\}\right\}$

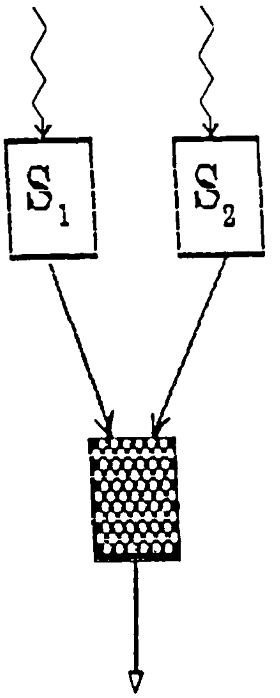
Figure 1c. $\mathfrak{R}(2)\left\{S_3\otimes\mathfrak{R}(2)\{S_2\otimes\overline{\mathfrak{R}(2)S}_1\}\right\}$

Figure 1d.

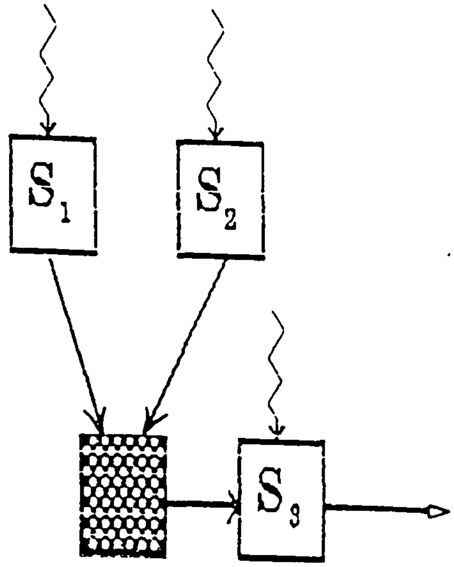
$$\mathfrak{R}(2)\left\{\overline{\mathfrak{R}(2)\left\{S_3\otimes\mathfrak{R}(2)\{S_2\otimes\overline{\mathfrak{R}(2)S}_1\}\right\}}\otimes\left\{\mathfrak{R}(2)\{S_6\otimes\mathfrak{R}(2)\{\overline{\mathfrak{R}(2)S}_4\otimes\overline{\mathfrak{R}(2)S}_5\}\}\right\}\right\}$$

The interior restrictions $\mathfrak{R}(2)$ represent a limitation of the communication channels to two-fold signals. The final overall restriction $\mathfrak{R}(2)$ represents the fact that there are only two courses of action available. The notation is easily generalized to admit other capacities. Note that this description of a sensor or network, and the concept of a doc (like the concept of an ROC) does not specify a particular tuning of the sensor. Similarly, the restriction operator represents the whole range of possible choices for the restriction.

Figure 1a. An example of "fusion" structure. Corresponding to the expression $\mathfrak{R}(2)S_i$ there is a box representing sensor S_i, from which there comes an arrow representing a two-fold message. The wavy line represents a signal $y \in Y_i$. The solid box represents the sensor product of the two restricted sensors.



A



B

Figure 1b. The fusion of messages from sensors 1 and 2 is combined in sensor product with the full information from sensor 3, forming a series structure.

Figure 1c. A pure series structure.

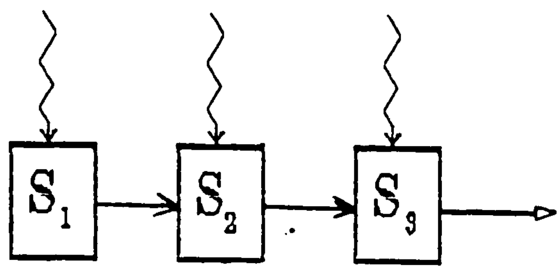
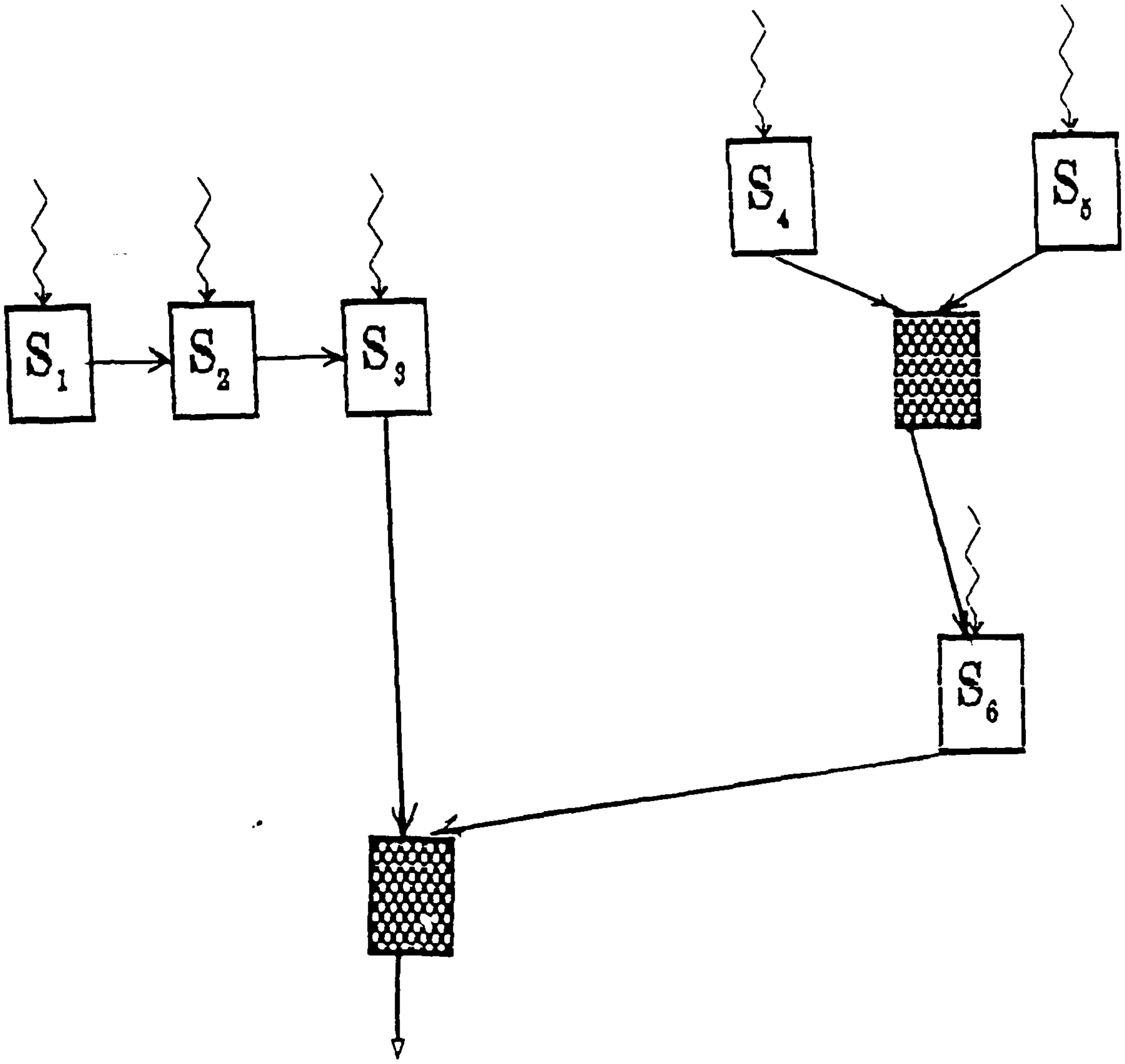


Figure 1d. A more complicated structure combining the features of Figures 1b and 1c through a final fusion.



In Part II we discuss how the optimal decision process for a set of alternative hypotheses $\mathbb{H}$ and a set of possible actions $\mathbb{A}=\{1,2,...\ A\}$ requires only the doc characterizing the network as a whole. We show how problems with "complete knowledge" of the cost matrix $C(a,h)$ and of the prior probabilities $p_1,...p_H$ are solved using the doc. We also describe the "Neyman-Pearson" problem (NP), in which C and p are not needed, and show that it is solved by the boundary of the doc.

This fact, that the doc solves the NP and all possible Bayesian problems, is very important, because the doc $\mathfrak{D}$ is completely determined by the hardware. It is a concrete engineering characterization of the network. The costs, $C(a,h)$, and the prior probabilities $p_1,...p_H$ are likely to be much softer. They do not originate in engineering constraints, and may change rapidly.

1.5 Organization of this paper. Acknowledgments.

The organization of this paper is as follows. Section 2 contains some examples of the doc $\mathfrak{D}$(S), the full product $\otimes$ and the k-fold restriction $\mathfrak{R}_k$ for cases in which the space of signals, $\mathbf{Y}$, is discrete. Section 3 gives examples for continuous signal sets. Section 4 presents fundamental network considerations for the specific case H=2. Section 5 gives some specific results for this case, including an example of spontaneous symmetry breaking, an example of a non-convex doc, a counter-example for series structure and the optimization procedures for any fixed topology, with either free or fixed combinative logic.

Part II begins with Section 6, which covers the use of the doc to solve both the Neyman-Pearson and the Bayesian problems. Section 7 contains a discussion of the discontinuities of system tuning parameters in the case of fusion, and the continuity of the best achievable cost. Section 8 shows how resource constraints modify the decision process, and may lead to a cost (performance gap) when only deterministic tunings are available. Section 9 illustrates the application of the doc and its boundary to the problem of team action.

The value of the doc lies in the fact that it permits a value-free comparison of alternative architectures, however complicated.

2. Some discrete examples of the detector operating characteristic $\mathfrak{D}(S)$.

Quite generally, the signal set and conditional probability distributions which define a sensor can be described by a fundamental table of numbers.

$$\begin{array}{lcccccccc} y\epsilon \mathbf{Y} & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ f_{h=1}(y) & .1 & .3 & \ldots \\ f_{h=2}(y) & .4 & .0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ f_{h=H}(y) & .1 & .2 & \ldots \end{array} \qquad (1)$$

The columns of the table are labeled by the elements of the signal set **Y**, which is taken, throughout this section, to be discrete. Even when the physical reality is a continuous signal, the practicalities of measurement will always force us to assign the observations to a finite number of discrete bins. With this in mind we will often refer to the elements of the signal set as "bins." The elements in each row represent the conditional probability that the observed signal will have the indicated value, provided that the state of nature which labels the row is indeed true. The elements in a row are called the "values of the response function corresponding to the indicated state of nature." The row sums are 1, and the column sums have no particular meaning. We will restrict our examples to the case in which the number of possible states of the world H=2. We bear in mind, however, that the common usage of "0" and "1" as the labels suggests an asymmetry among the hypotheses and the actions which does not exist in general. (Although there will be at least one action which is "best" if a given state of nature prevails, the remaining actions may be "wrong" to differing degrees, and their ordering will change according to which state of nature does indeed prevail.)

2.1 Specific Examples

Consider a specific concrete example S given by the fundamental table:

$$S_1 = \begin{bmatrix} y & = & 1 & 2 & 3 \\ f_1 & = & .6 & .3 & .1 \\ f_0 & = & .1 & .3 & .6 \end{bmatrix} \qquad (2)$$

For any sensor, the notion of "tuning" amounts to specifying the circumstances under which a particular action (or signal, if the sensor is imbedded in a network) will be chosen. For example, we may represent the set of signals leading to the action "a=1" as Y(a=1), which is a subset of **Y**. The corresponding probabilities to act, given the alternative states of nature "0" and "1" are represented in the table of bin combinations:

$$\mathfrak{D}(S) = \left| \begin{array}{ccccccccc} Y(a=1): & \{\emptyset\} & \{1\} & \{2\} & \{3\} & \{1,2\} & \{1,3\} & (2,3\} & \{1,2,3\} \\ h=1 & .0 & .6 & .3 & .1 & .9 & .7 & .4 & 1.0 \\ h=0 & .0 & .1 & .3 & .6 & .4 & .7 & .9 & 1.0 \end{array} \right| \tag{3}$$

The pair of conditional probabilities in any given column may be taken as the coordinates of a point in a two-dimensional space. The dimensionality of the space is given not by the number of actions, but by the number of hypotheses (H). We refer to this set of points in an abstract space as the doc $\mathfrak{D}(S)$. It contains all of the useful information in the table. The points in the doc may each be labeled by the subsets to which they correspond.

We give two other examples to solidify the concept. Consider first the "broken detector." A broken detector always gives the same signal, which we choose to be "3". In talking about the case of only two actions, to which we now restrict ourselves, it is convenient to describe "a=1" as "act" and "a=2" as "do nothing."

The fundamental table of the sensor is:

$$S_{BROKEN} = \begin{bmatrix} y & = & 1 & 2 & 3 \\ f_1 & = & .0 & .0 & 1.0 \\ f_0 & = & .0 & .0 & 1.0 \end{bmatrix} \tag{4}$$

The table of the doc becomes:

$$\mathfrak{D}(S)=\left|\begin{array}{ccccccccc} Y(a=1): & \{\emptyset\} & \{1\} & \{2\} & \{3\} & \{1,2\} & \{1,3\} & \{2,3\} & \{1,2,3\} \\ h=1 & .0 & .0 & .0 & 1. & .0 & 1. & 1. & 1. \\ h=0 & .0 & .0 & .0 & 1. & .0 & 1. & 1. & 1. \end{array}\right| \tag{5}$$

There are only two distinct points in the doc. One point corresponds to all tunings in which the set Y(a=1) contains the element "3" of the signal set. With this tuning we will "always act." The other point corresponds to all other subsets of Y, and with this tuning we will "never act."

2.2 Sensor Product S⊗T

Consider a second detector whose bins are not necessarily the same as those of S_1, with fundamental table:

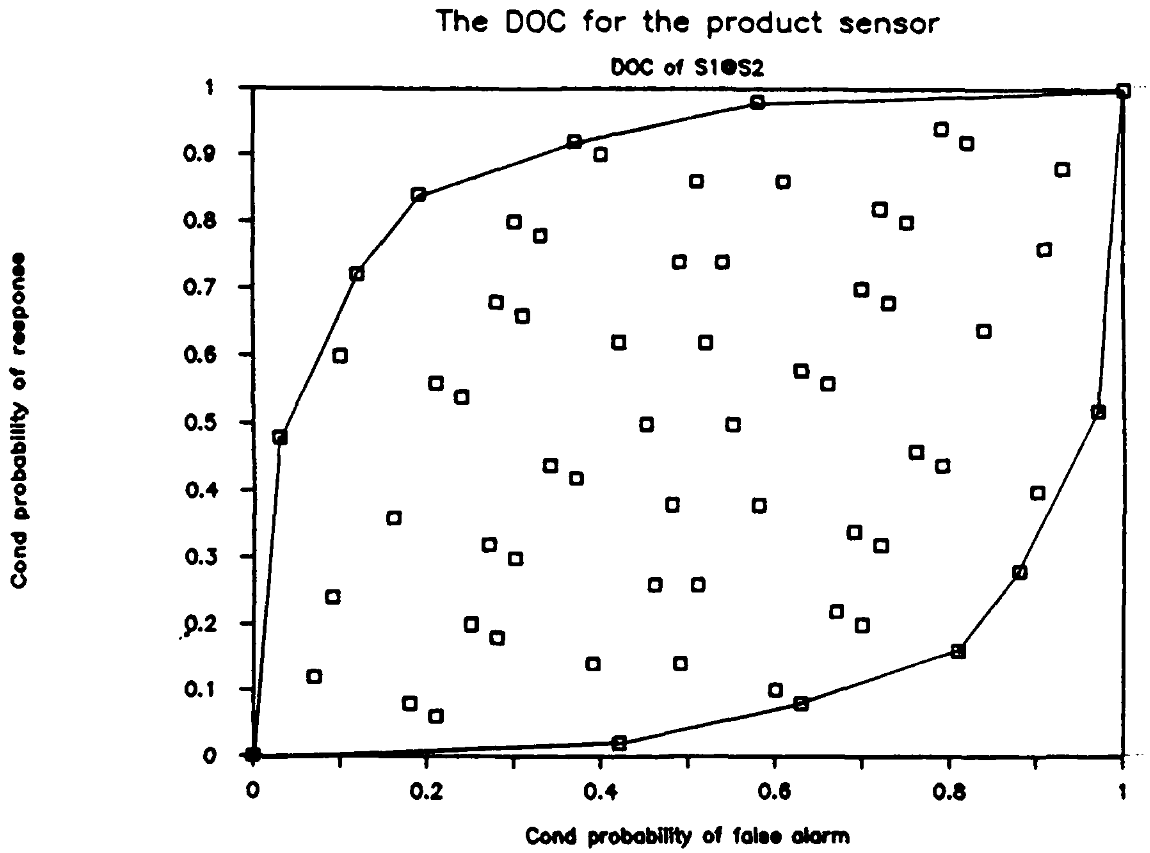
$$S_2=\begin{bmatrix} 5 & 6 \\ .8 & .2 \\ .3 & .7 \end{bmatrix} \tag{6}$$

We define the full product sensor, $S_1 \otimes S_2$ by the product table. It represents all of the information that can be given by the two sensors together, and has six bins which may be labeled as 15, 16, 25, 26, 35 and 36. Quite generally, the rows of the fundamental table will be the conditional joint probability distributions. If the two sensors are (stochastically) independent the table for the full product is determined by the two individual tables. Specifically we have:

$$S_3= S_1 \otimes S_2|_{indep} = \begin{bmatrix} 15 & 16 & 25 & 26 & 35 & 36 \\ .48 & .12 & .24 & .06 & .08 & .02 \\ .03 & .07 & .09 & .21 & .18 & .42 \end{bmatrix} \tag{7}$$

In what follows we will assume stochastic independence throughout. Since there are 2x3=6

Figure 2. The detector operating characteristic (doc) $\mathfrak{D}$ for the product of two discrete sensors. Each small square represents an achievable value of the two conditional probabilities. The boundary $\mathfrak{B}^+$ plays a central role in decision analysis. From the boundary one may reconstruct the fundamental table of the product sensor.



bins in the product detector, there are $2^6=64$ points in the table of bin combinations which defines the doc. The resulting doc $\mathfrak{D}(S_3)$ is shown in Figure 2. The fundamental table of the sensor may be reconstructed from the difference vectors formed along the upper boundary, which are suggested by the light line in Figure 2. This line is, in fact, the ROC for this system, except that, since the system and its doc are discrete, only the vertices are realizable in a deterministic system.

Note that the doc of a sensor product must always contain the doc of either of the factors because one sensor may be tuned to the point (1,1), in which case the products are all possible tunings of the other sensor. The product procedure can be followed in the case of a continuous signal set, by binning to any desired degree of approximation, in order to produce a standard discrete representation of the doc of a continuous system.

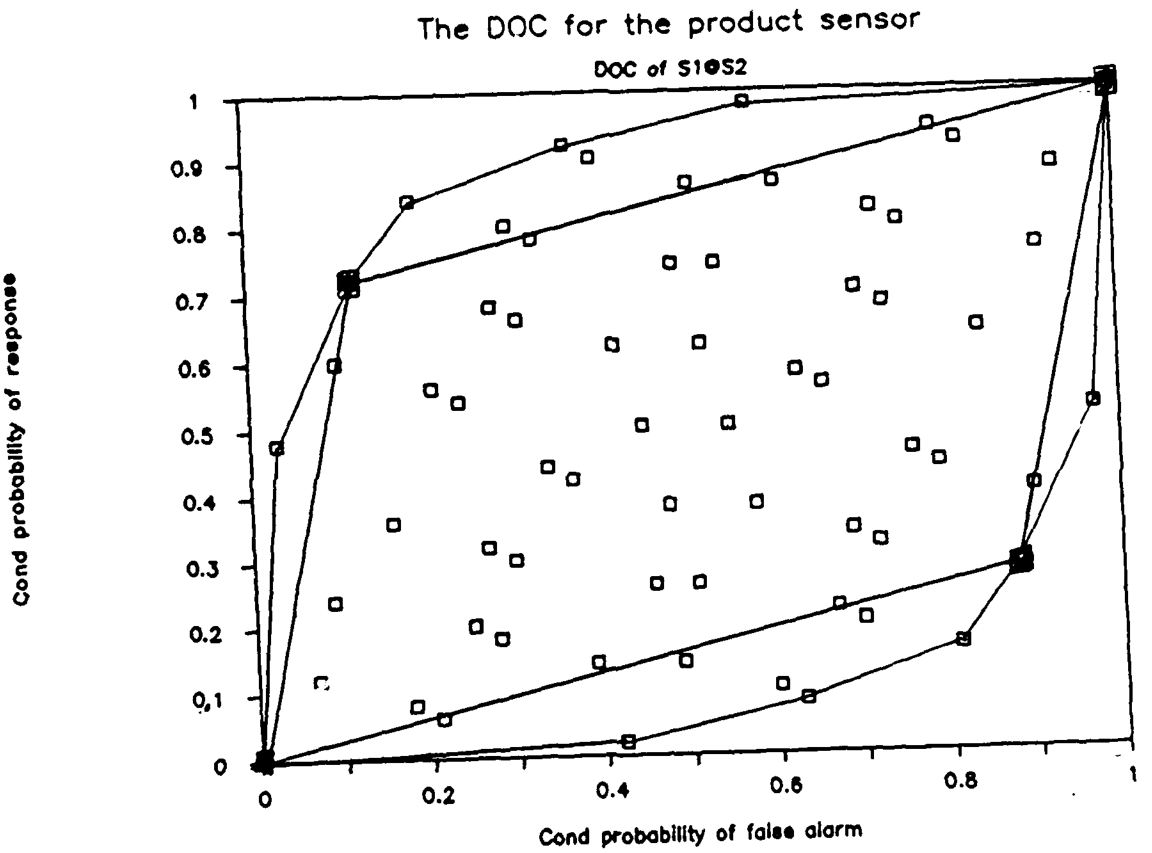
2.3 Restriction of Sensors $\mathfrak{R}(M)S$

When a sensor communicates through a finite network channel it may not be able to pass on all of the available information. It must code the observed signal $y \epsilon \mathbf{Y}$ into some M-fold message. To do this the signal set $\mathbf{Y}$ is decomposed into a union of non-overlapping subsets $\mathbf{Y}(m=1)$, $\mathbf{Y}(m=2)$, ...$\mathbf{Y}(m=M)$. Each such partition represents a "tuning" of the sensor. In general the channel capacity (M) is less than the total number of bins. For example, with M=2, and the sensor $S_1 \otimes S_2$ there are 64 tunings, corresponding to the points in the doc. Among them is the totally uninformative choice $\mathbf{Y}(m=1)=\{25, 26\}$, which lies on the principal diagonal of the unit square containing the doc. There are also 5 maximally informative possibilities:

$\mathbf{Y}(m=1)=\{15\}$, $\{15,25\}$, $\{15,25,16\}$, $\{15,25,16,35\}$, and $\{15,25,16,35,26\}$.

Further discussion is postponed to Part II. We represent any one of the possible tunings by the general expression $\overline{\mathfrak{R}(2)S_3}$. For example, one specific tuning is: $\overline{\mathfrak{R}(2:\{15,25\})S_3}$, whose doc is shown in Figure 3.

Figure 3. A restriction of the sensor of Figure 2. The points of the original doc are shown for reference. The doc of the restricted sensor, which has only two messages, consists of the four solid squares.



3. Some continuous examples.

We have already remarked that it is practically necessary to replace a quantity that is "in principle" continuous by a finite set of discrete bins. In other cases it is convenient to approximate something that is fundamentally discrete as being essentially continuous. We now consider the extension of the concepts of the doc $\mathfrak{D}(S)$, the full product $\otimes$, and the n-fold restriction $\mathfrak{R}(n)$ to cases in which the signal set **Y** is continuous.

3.1 The case of exponential response functions.

The case of exponential response functions is particularly tractable, and will be pursued until its simplicity proves to be an embarrassment. The signal set $\mathbf{Y}=[0,\infty]$. The response functions are $f_1(y)=e^{-y}$ and $f_0(y)=ne^{-ny}$. The set of points in the doc is precisely the allowed region of our previous paper [Blankenbecler and Kantor88]. Referring to the previous section, we see that even in the discrete case the set of points in the doc quickly becomes very dense. We may readily find the boundary (in the case of only two hypotheses) by ordering the points of **Y** in decreasing order of the ratio $f_1(y)/f_0(y)$. This provides a parametric representation of the boundary of the doc, which is sufficient for further numerical calculation.

If we call the parameter involved "z," a suitable choice is given by $\mathbf{Y}(m=1;z)=[1/z,\infty]$. Using "F,D" to represent points on the boundary we have at once:

$$F_0(\mathbf{Y}(m=1;z))=n\int_{1/z}^{\infty} e^{-ny}\,dy = e^{-n/z} \tag{8}$$

and

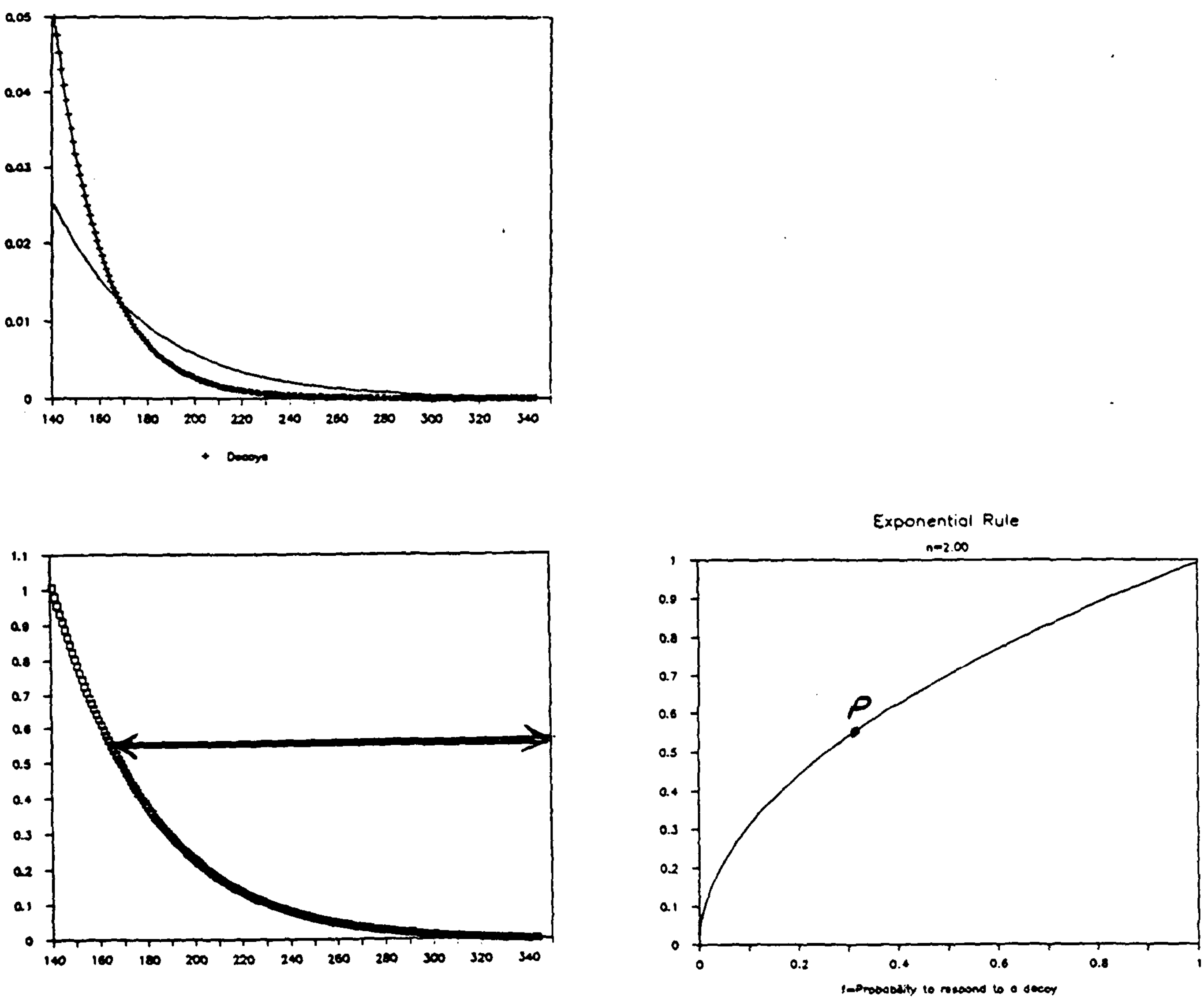
$$F_1(\mathbf{Y}(m=1;z))=\int_{1/z}^{\infty} e^{-y}\,dy=e^{-1/z}. \tag{9}$$

In this case a simple analytic relation describes the upper boundary of the doc:

$$F_1(F_0) = F_0^{\,n} \qquad (n\geq 1) \tag{10}$$

We may represent all of the salient features of this problem in a family of three related graphs, as shown in Figure 4. The first figure shows the response functions on any convenient scale, as a function of y, or a transformed label. The third figure shows the upper boundary

Figure 4. Aspects of a sensor with continuous signal set. The response function, the upper boundary of the doc, and the map from the boundary to trigger regions in the signal set are shown for the exponential model of the text.



$\mathfrak{B}^+$ of the doc $\mathfrak{D}(.)$. The lower boundary is determined by the symmetry of the doc under the transformation $F \rightarrow (1-F)$ and $D \rightarrow (1-D)$, the relabeling of actions. We adopt here the convenient notations:

$$F = F_0(\mathbf{Y}(z)) \tag{11}$$

$$D = F_1(\mathbf{Y}(z)) \tag{12}$$

corresponding to the notion that f_0 represents a "false alarm," while f_1 represents a "true detection event." The middle part of the figure shows the translation of any particular operating point on the boundary of the doc into a corresponding "trigger region" $\mathbf{Y}(z)$. The parameter z itself need never be made explicit.

3.2 Equivalent Sensors

Corresponding to the fact that, for a discrete system, permuting the columns of the fundamental table does not change the doc, there are an infinity of transformations of the response functions which will leave the doc unchanged in the continuous signal case. A simple example is given by the Rayleigh distributions:

$$f_1(w) = 2nwe^{-nw^2} \tag{13}$$

and

$$f_0(w) = 2we^{-w^2}. \tag{14}$$

The transformation $y = w^2$; $dy = 2wdw$ shows that these two sets of response functions, the Rayleigh and the exponential, have exactly the same integrals over corresponding regions in their respective signal sets, and hence will have the same doc.

There is very little difficulty in principle in extending the parametric representation to any computable forms for f_0 and f_1. A more complex example, on the signal set $\mathbf{Y}=[-\infty,\infty]$ is:

$$f_1(y) = \frac{1}{20\sqrt{2\pi}}\left[(1/3)e^{-(y-200)^2/2(20)^2} + (1/3)e^{-(y-250)^2/2(20)^2} + (1/3)e^{-(y-300)^2/2(20)^2}\right] \tag{15}$$

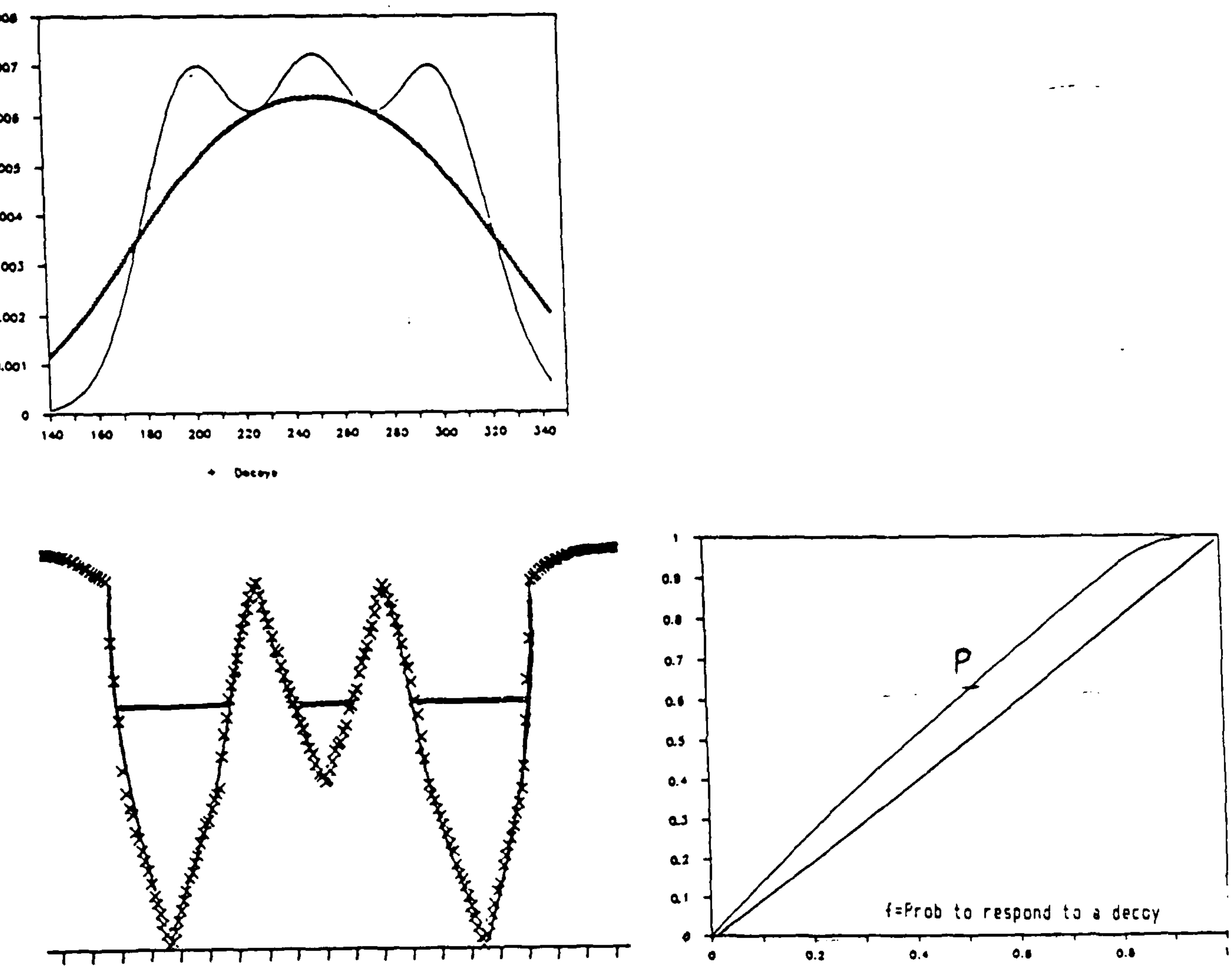
and

$$f_2(y)=\frac{1}{40\sqrt{2\pi}}\left[(1/3)e^{-(y-200)^2/2(40)^2}+(1/3)e^{-(y-250)^2/2(40)^2}+(1/3)e^{-(y-300)^2/2(40)^2}\right] \tag{16}$$

The doc has been calculated numerically by ordering unit bins centered at $y=150,\ldots 350$, as described above. It is shown in Figure 5. Note that the trigger regions may be quite complex, because the response functions do not have a monotone likelihood property with respect to the label or variable "y".

It may be shown that, provided the ratios of the response function do not vary too rapidly, the doc $\mathfrak{D}$ corresponding to any fundamental table whose signal space is continuous will be a convex set. That is, there will not be any isolated extreme points, or holes within the boundary $\mathfrak{B}$. This property is useful in studying the fundamental operations of the sensor calculus.

Figure 5. A complex sensor with continuous signal set. The response functions, upper boundary of the doc (the ROC) and the map to trigger regions are shown for compound overlapping Gaussians. The solid line on the lower left graph is the trigger region corresponding to the point $P \in \mathfrak{B}^+$.



3.3 Standard Forms for Sensor Tables

We have mentioned that different sensors may have the same doc. It is therefore useful to introduce the notion of a standard doc. This may be done in two ways. One leads to a continuous signal set, and may be useful for conceptual purposes. The other leads to a discrete signal set, which is essential for calculation. We suppose that the upper boundary $\mathfrak{B}^+$ of the doc is given in the form $D(F)$. The derivative $D'(F)$ may be shown to exist, from the right, for all $F<1$.

Continuous Standard Form:

$$\mathbf{Y}=[-D(0),1] \tag{17}$$

$$f_0(y)=\begin{cases} 0 & y<0 \\ 1 & 0\leq y\leq 1 \end{cases} \tag{18}$$

and

$$f_1(y)=\begin{cases} 1 & -D(0)\leq y\leq 0 \\ D'(y) & 0<y\leq 1 \end{cases} \tag{19}$$

Discrete Standard Form (N points):

Define the auxiliary function:

$$f_{aux}(\theta)=\min_{D'(F)\leq \tan\theta} F. \tag{20}$$

For n=1 to N:

$$F_n=f_{aux}(n\pi/(2N))-f_{aux}((n\text{-}1)\pi/(2N)). \tag{21}$$

$$D_n=D\left\{f_{aux}(n\pi/(2N))\right\}-D\left\{f_{aux}((n\text{-}1)\pi/(2N))\right\}. \tag{22}$$

This construction divides the continuous interval from 0 to 1 into portions over which the slope of the boundary of the doc changes by a fixed amount, $\pi/2N$. This construction makes use of the convexity of the doc and, hence, the fact that its boundary has a monotonically changing

slope. [Technically, as we draw the doc, the upper boundary is concave, and the lower boundary is convex.]

3.4 The Full Sensor Product:

As a simple example we form the full sensor product of two identical sensors with exponential response. This is the same as having all the information from two such sensors before making a decision. Or, it can be regarded as making two successive (independent) observations with the same instrument, before reaching a decison.

We introduce the transparent notation:

$$S=\begin{bmatrix} \mathbf{Y}=[0,\infty] \\ d=e^{-y} \\ f=ne^{-ny} \end{bmatrix} \tag{23}$$

to represent the sensor with $\mathbf{Y}=[0,\infty]$ and with the response functions indicated. We see that the full sensor product has the representation:

$$S\otimes S=\begin{bmatrix} [0,\infty]x[0,\infty] \\ d(x,y)=e^{-(x+y)} \\ f(x,y)=n^2 e^{-n(x+y)} \end{bmatrix} \tag{24}$$

The obvious parametrization for determination of the boundary of the doc is $t=x+y$. The corresponding trigger regions are of the form $\mathbf{Y}(t_0)=[t_0,\infty]$. The element of integration becomes tdt, with the results:

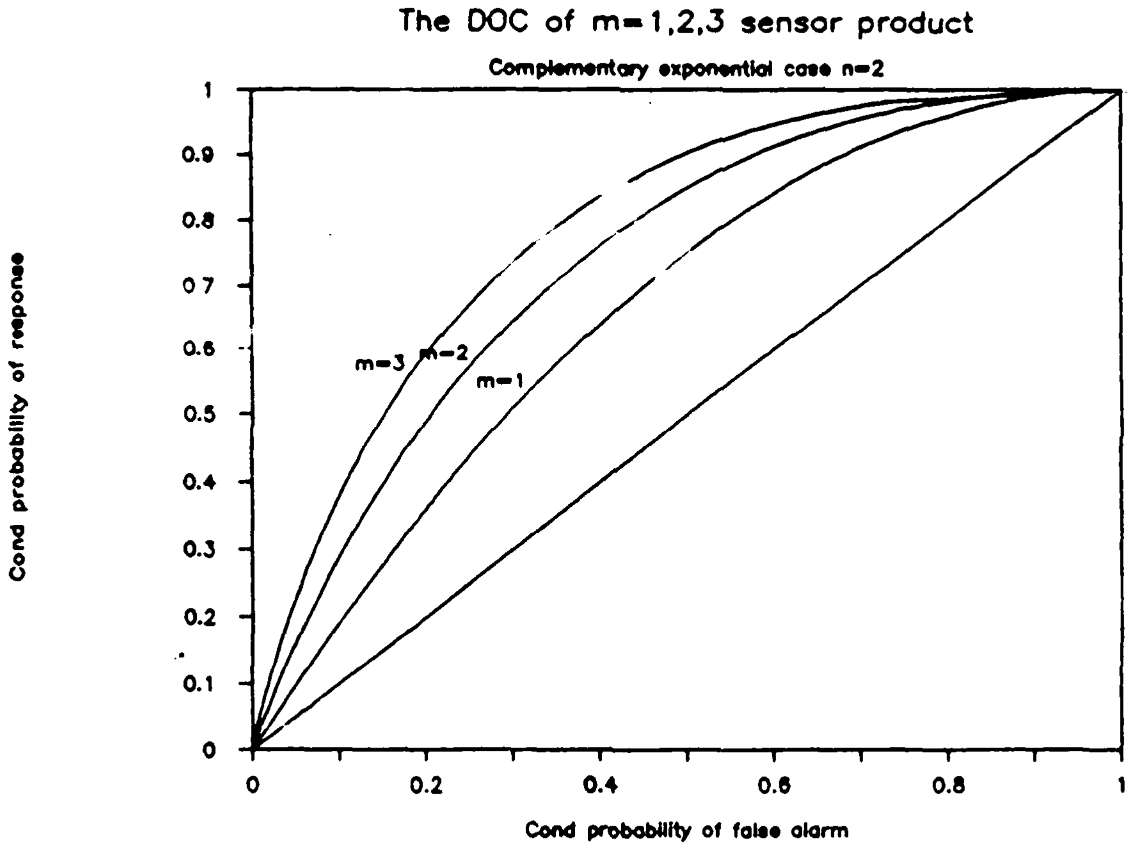
$$D(t)=(1+t)e^{-t} \tag{25}$$

$$F(t)=(1+nt)e^{-nt} \tag{26}$$

A general recursive formula is given by:

$$D_k(t)=D_{k-1}(t)+\frac{t^{k-1}}{(k-1)!}e^{-t} \qquad k=1,2,\ldots \tag{27}$$

Figure 6. The full sensor product. The upper boundary of the doc, $\mathfrak{B}^+$ is shown for the two and three-fold product of the exponential sensor with itself. Note that there are diminishing returns in the continued improvement that repeated measurement represents.



$$F_k(t) = F_{k-1}(t) + \frac{n^{k-1} t^{k-1}}{(k-1)!} e^{-nt} \qquad k=1,2,\ldots \tag{28}$$

with:

$$D_0(t) = F_0(t) = 0. \tag{29}$$

Although we generally cannot express $D(F)$ in closed form, there is no difficulty in preparing graphs, or performing further calculations on the basis of these formulae and results. Examples showing the upper boundary $\mathfrak{B}^+$ of the doc for the two-fold and three-fold sensor product for the exponential case are shown in Figure 6.

It is reasonable to suppose that if two sensors have the same doc, the doc of their product with other sensors will not depend on the specific representation $\left\{\Upsilon, d(y), f(y)\right\}$ that is used. The reader may verify this by repeating the preceeding calculation, replacing either or both of the sensor descriptions by the Rayleigh form.

4. Fundamental Network Elements

The fundamental operations that go to build up a network have already been defined: the communication restriction $\overline{\mathfrak{R}(M)S}$ and the full sensor product $S \otimes S$. There are some basic topologies which it is instructive to examine in detail, both to illustrate the calculational techniques, and to sharpen our intuition.

4.1 Specific and General Restrictions $\mathfrak{R}(M,t)S$ and $\mathfrak{R}(M)S$

We recall that $S \otimes T$ produces a new composite sensor, with its associated doc, which is always (in a sense to be made clear in Section 6) at least as good as either S or T. On the other hand, for any particular tuning t, $\overline{\mathfrak{R}(M,t)S}$ is a sensor which is, in general, not as good as S, because it has only an M-fold output. The tuning t determines the meaning of that output. It is logically equivalent to a decomposition of the signal set into a set of nonoverlapping subsets Y(m=1),...,Y(m=k), but, in practice, t may be represented in a variety of ways. In particular, when we want to determine the boundary $\mathfrak{B}$ of a compound doc $\mathfrak{D}(S \otimes T)$, we need only consider extreme points of the constituent docs. When their signal sets Y(S) and Y(T) are suitably continuous the boundary is a continuous set, containing all the extreme points. When the signal sets are discrete, the "boundary" is the set of extreme points.

Note that the simpler expression $\overline{\mathfrak{R}(2)S}$ represents a more complex object than $\overline{\mathfrak{R}(2,t)S}$ since it is a set:

$$\overline{\mathfrak{R}(2)S} = \left\{ \overline{\mathfrak{R}(2,t)S} : \text{t a possible tuning of S} \right\} \tag{30}$$

.milarly:

$$\overline{\mathfrak{R}(k)S} \otimes \overline{\mathfrak{R}(l)T} = \left\{ s \otimes t\text{: } s \epsilon \overline{\mathfrak{R}(k)S},\ t \epsilon \overline{\mathfrak{R}(l)T} \right\}. \tag{31}$$

Thus the elements of the set $\overline{\mathfrak{R}(k)S} \otimes \overline{\mathfrak{R}(l)T}$ are labelled by two tunings: one for S and one for T. In the examples of this paper those tunings are represented by real numbers in the unit interval, corresponding to the probability of a "false alarm." This representation of the tuning is possible when there are only two hypotheses and only two actions or messages.

Any element $s \otimes t \ \epsilon \ \overline{\mathfrak{R}(k)S} \otimes \overline{\mathfrak{R}(l)T}$ is itself a sensor, and it has a signal set with $k \times l$ points, which are labelled by the messages coming from S and T, under the tunings selected.

When an action a=1,2,...,A is to be selected from a set A the compound sensor $s\otimes t$ must itself be tuned. Such a tuning is a decomposition of $\mathbf{Y}(s\otimes t)$ into A non-overlapping subsets. In the examples of this paper A=2 and the decomposition is specified by giving the set $\mathbf{Y}(a=1)$ which, by complementation, specifies the set $\mathbf{Y}(a=2)$.

4.2 Binary Messages

In the special case of binary messages $m\epsilon\{0,1\}$ it is natural to call the tuning of $s\otimes t$ a LOGIC. The signal set is $\mathbf{Y}(s\otimes t)=\{00,01,10,11\}$. There are $2^4-1=15$ non-empty subsets in the doc $\mathfrak{D}(s\otimes t)$. Each such subset corresponds to a logical expression. For example {01,10} corresponds to [m(s)=1 or m(t)=1 but not both], which could be expressed as the exclusive or: XOR(s,t).

Corresponding to a given logic there is a fundamental polynomial which appears in each row of the table characterizing the sensor system. It is defined in terms of the binary patterns $m=(m_1,\cdots m_2)$ appearing in the logic:

$$Q_{LOGIC}(x_1,\cdots,x_n)=\sum_{m\epsilon LOGIC}\prod_{i=1}^{n} x_i^{m_i}(1-x_i)^{(1-m_i)} \tag{32}$$

Here n is the number of independent sensors for which a fusion center has been used. This particular form depends both on the fact that only two complementary trigger sets arise at any sensor (because of the two-fold messages), and the fact that 1 can be written as the power x^0.

There is an important set of inequalities restricting the LOGICS that can be extreme points of $\mathfrak{D}(s\otimes t)$. We recall that the elements of the table defining $s\otimes t$ are products for stochastically independent sensors. In the special case of two-fold messages we may simplify the notation, using:

$$d_s=d(\mathbf{Y}(m(s)=1)=Prob(m(s)=1 \text{ given } h=1) \tag{33}$$

$$f_s=f(\mathbf{Y}(m(s)=1)=Prob(m(s)=1 \text{ given } h=0), \tag{34}$$

with similar expression for the sensor t.

We also set:

$$\bar{x} \equiv 1 - x. \tag{35}$$

Then the fundamental table describing s⊗t has as its columns all triples of the form:

$$\begin{bmatrix} & \mathrm{m(s)m(t)} & & \\ \mathrm{d_s}^{\mathrm{m_s}} & \overline{\mathrm{d_s}}^{\overline{\mathrm{m_s}}} & \mathrm{d_t}^{\mathrm{m_t}} & \overline{\mathrm{d_t}}^{\overline{\mathrm{m_t}}} \\ \mathrm{f_s}^{\mathrm{m_s}} & \overline{\mathrm{f_s}}^{\overline{\mathrm{m_s}}} & \mathrm{f_t}^{\mathrm{m_t}} & \overline{\mathrm{f_t}}^{\overline{\mathrm{m_t}}} \end{bmatrix} \tag{36}$$

The doc $\mathfrak{D}$ has elements corresponding to all possible sums of these expressions.

Without loss of generality we may suppose that $\mathbf{Y}(\mathrm{m}=1)$ is chosen so that

$$\frac{\mathrm{d_s}}{\mathrm{f_s}} \geq \frac{1-\mathrm{d_s}}{1-\mathrm{f_s}} = \frac{\overline{\mathrm{d_s}}}{\overline{\mathrm{f_s}}}. \tag{37}$$

with a similar relation for $\mathbf{Y}(\mathrm{m(t)}=1)$.

It is easy to see that if a "trigger region" contains the point m(s)m(t)...m(n) and does not contain all the points $\mathrm{m}'(\mathrm{s})\mathrm{m}'(\mathrm{t})\ldots\mathrm{m}'(\mathrm{n})$ for which any $\mathrm{m}' \geq$ the corresponding m, then it is not an extreme point of the upper boundary $\mathfrak{B}^+$ doc of the composite sensor.

We sketch the proof for the case of two-fold signals. A corresponding result may be proven for k-fold signals in the same way. The statement is empty if the region is {11}. Otherwise, suppose that for sensor t, the trigger region contains some point with $m(t)=0$. We need only show that the corresponding point in the doc is inside the convex hull of the doc. Without loss of generality we need only consider cases lying in the triangle $d \geq f$. We decompose the given point into:

$$(f,d) = (f_A + f_0,\ d_A + d_0) \tag{38}$$

where $(f_0,\ d_0)$ is the vector in the doc space corresponding to the point with $m(t)=0$ in the signal set. Let $(f_1,\ d_1)$ represent the vector in doc space corresponding to the same point in the signal set, but with $m(t)=1$. We show that (f,d) is not an extreme point of the doc $\mathfrak{D}$ by showing that is lies below the line joining the points $\mathrm{A} \equiv (f_A, d_A)$ and

$B \equiv (f_A + f_0 + f_1, d_A + d_0 + d_1)$. To prove this we note that the f-coordinate of $(f_1/(f_0+f_1))A + (f_0/(f_0+f_1))B$ is f, while the d-coordinate is:

$$d_A + \frac{f_0}{f_0+f_1}(d_0 + d_1) \tag{39}$$

$$= d_A + \frac{f_0}{f_0+f_1}(d_o + f_1 d_1/f_1) \tag{40}$$

$$\geq d_A + \frac{f_0}{f_0+f_1}(d_o + f_1 d_0/f_0) \tag{41}$$

$$= d_A + d_0 = d. \tag{42}$$

The set of LOGICS is further reduced by the observation that every sensor must play a role. For example, the doc of any logic that is independent of the message from sensor S is contained within the doc formed by multiplying the corresponding polynomial Q_{LOGIC} by ${x_s}^{m_s}$. For, we could freeze sensor S to the "ON" position and recover Q_{LOGIC}. When a logic is independent of the message from sensor S it involves x_S only in the form $x_S + \overline{x_S} = 1$. For example, with two sensors the logic {10,11} corresponds to $Q(x) = x_s \overline{x_t} + x_s x_t = x_s$. Its doc is contained within the doc of either {11}=Y(AND), or {01,10,11}=Y(OR).

Of course the degenerate cases $\mathbf{Y} = \{\emptyset\}$ and $\mathbf{Y} = \{00,01,10,11\}$ are equivalent to having no detector at all, and need not be considered. They correspond to the BROKEN doc: $\mathfrak{D} = \{(0,0),(1,1)\}$.

These three principles are the only ones that we know for reducing the set of possible logics. Detailed examples are given in [Cherikh88; Thesis CWRU] where the relation between these rules and a Lagrange multiplier formalism for definition of the boundary of the doc is developed.

4.3 Specific Descriptions of 2-fold and 3-fold fusion.

Using these rules we find that the cases to be considered are:

(i) Product combination of two sensors: $\overline{\mathfrak{R}(2)S}_1 \otimes \mathfrak{R}(2)S_2$:

$$\text{LOGIC} = \{11\} \;\; \text{AND}$$
$$= \{01,10,11\} \;\; \text{OR} \tag{43}$$

(ii) Product combination of three sensors: $\overline{\mathfrak{R}(2)S}_1 \otimes \overline{\mathfrak{R}(2)S}_2 \otimes \overline{\mathfrak{R}(2)S}_3$:

$$\text{LOGIC} = \{111\} \;\; \text{AND}$$
$$= \{011,101,110\} \;\; \text{MAJORITY RULE or 2 OUT-OF 3}$$
$$= \{011,100,101,110,111\} \;\; \text{1 OR (2 AND 3)}$$

plus two cyclic permutations.

$$= \{101,110,111\} \;\; \text{1 AND (2 OR 3)}$$

plus two cyclic permutations.

$$= \{001,010,011,100,101,110,111\} \;\; \text{OR.} \tag{44}$$

The specific computations needed to trace the extreme points of the doc can always be written as:

$$D(F) = \max_{Q_{\text{LOGIC}}(f_s, f_t) \le F} Q_{\text{LOGIC}}(d_s, d_t). \tag{45}$$

Since the function $D_r(F_r)$ is monotonically increasing for $r = s$ or t, the weak inequality constraint may always be replaced by equality if the doc set is continuous. Thus, for two detectors in this "fusion" situation the optimization involves a search over one variable, For three detectors it involves a search over two variables.

4.4 Series Structures:

When one of two sensors is directly accessible, and the other is only accessible over a k-

fold channel we call the structure "series." The corresponding doc is represented by $\overline{\mathfrak{R}(k)S_1} \otimes S_2$. We say that sensor S_2 is "downstream" from sensor S_1. Each point in the doc of the combined system is achievable as the product of one or more pairs of points in the doc of S_2 and some member of $\overline{\mathfrak{R}(2)S_1}$. Let us examine the structure of such pairs.

For any particular choice of the "tuning" of S_1 the doc of $\overline{\mathfrak{R}(2)S_1}$ is a set of four points: (0,0), $(f_1,d_1(f_1))$ and their reflections under $(f,d) \rightarrow (1\text{-}f,1\text{-}d)$. The only non-trivial choice of a trigger set is $\mathbf{Y}(m{=}1) \rightarrow (f_1,d_1)$ and $\mathbf{Y}(m{=}0) \rightarrow (1-f_1,1-d_1)$. The values of (f_1,d_1) can range over the entire $\mathfrak{D}(S_1)$. Similarly, the points in the doc of S_2 can range over $\mathfrak{D}(S_2)$.

On the one hand, the doc $\mathfrak{D}$ of the series case is a restriction of the doc $\mathfrak{D}$ for the full sensor product. We consider first the discrete case, with both signal sets having 3 elements. $\mathbf{Y}_1=\{1,2,3\}$ and $\mathbf{Y}_2=\{4,5,6\}$. The signal set of $S_1 \otimes S_2$ is {14,15,16,24,25,26,34,35,36}, which has 9 points. The doc will consist of $2^9=512$ points. The signal set of $\overline{\mathfrak{R}(2)S_1} \otimes S_2$ is somewhat more complicated. There are several possibilities for the signal from $\overline{\mathfrak{R}(2)S_1}$, depending upon the particular tuning, which we denote as $\overline{\mathfrak{R}(2;t)S}$. They are {∅,123}, {1,23}, {2,13}, {12,3} (4 possibilities in all, as their complements provide the remainder of the $2^3=8$ total range of possibilities.) However, these possibilities are not simultaneously available! When a specific tuning is made for the first sensor, one of these possibilities is available and the others are not. Thus there are only 2x3=6 elements in the $\mathbf{Y}$ of the series system, and not 9. Finally, if we are interested in finding the extreme points of the doc, one of the possible combinations, {2,13}, will not be of interest because it is not an extreme point in the doc of of S_1.

One way to visualize the relation between $\overline{\mathfrak{R}(2)S_1} \otimes S_2$ and $S_1 \otimes S_2$ is to form a table of the possible subsets of the product signal set. For the full sensor product, every element of the product set may be independently included in the trigger set. For the restricted product $\overline{\mathfrak{R}(2)S_1} \otimes S_2$, for every element in $\overline{\mathfrak{R}(2)S_1}$ the elements of S_2 may be chosen independently, and vice versa. But this means that the elements of S_1 must be assigned to two subsets once and for all, prior to the formation of the trigger set. Thus, the candidates to be on the boundary $\mathfrak{B}$ of the doc $\mathfrak{D}(\overline{\mathfrak{R}(2)S_1} \otimes S_2)$ correspond to the following points in the signal set $\mathbf{Y}$:
{14,15,16,234,235,236} or {124,125,126,34,35,36}

Since only one of these possibilities may be realized at a time, one could not choose the tuning {14,1245}. The upstream sensor cannot distinguish "2" from "1" and also from "3", because it communicates via a 2-fold channel.

4.5 Comparison of the series structure to fusion.

Although the series structure has a more limited signal set (and, hence, a restricted doc $\mathfrak{D}$) than the full sensor product, it is expected to be more general than the fusion structure $\overline{\mathfrak{R}(2)S} \otimes \overline{\mathfrak{R}(2)T}$. This may be verified by writing out explicitly the elements of the signal set for the 4 non-trivial possibilities of $(s \otimes t) \epsilon \overline{\mathfrak{R}(2)S} \otimes \overline{\mathfrak{R}(2)T}$. They are:

$$\begin{aligned}&\{14,156,234,2356\}\\&\{145,16,2345,236\}\\&\{124,1256,34,356\}\\&\{1245,126,345,36\}.\end{aligned} \tag{46}$$

The first two of these are contained within the first of the series possibilities. The remainder are contained within the second. Note that what appears as an elementary possibility in the fusion structure, such as "1245" (that is: S says 1 or 2 and T says 4 or 5) is a composite in the series structure, being the union of the elements 124 and 125.

4.6 Computation of a Series doc $\mathfrak{D}(\overline{\mathfrak{R}(2)S}_1 \otimes S_2)$ in the continuous case.

The series configuration can be thought of as a union of docs, corresponding to a set of sensor products:

$$\overline{\mathfrak{R}(2)S}_1 \otimes S_2 = \left\{ s \otimes S_2 \colon s \epsilon \mathfrak{R}(2,t) S_1 \text{ for some tuning } t \right\}. \tag{47}$$

The structure of the specific elements of the doc is a little tricky. In general, the elements of $s \epsilon \mathfrak{R}(2,t) S_1$ are of the form shown in Eq. 48:

$$\begin{bmatrix} m=1 & m=0 \\ D_1(t_1) & \overline{D_1(t_1)} \\ F_1(t_1) & \overline{F_1(t_1)} \end{bmatrix}. \tag{48}$$

That is, there are only two points, and each row has, as its coordinates, some point on the boundary $\mathfrak{B}(S_1)$ of the doc $\mathfrak{D}(S_1)$. The elements of $s \otimes S_2$ are sums of products with one factor drawn from this table and the other drawn from the table of S_2. We may write this sensor in general as a table (we suppress the row containing the labels):

$$\begin{bmatrix} D_1(t_1) & \overline{D_1(t_1)} \\ F_1(t_1) & \overline{F_1(t_1)} \end{bmatrix} \mathrm{X} \begin{bmatrix} d_2(y_1) & d_2(y_2) & \cdots \\ f_2(y_1) & f_2(y_2) & \cdots \end{bmatrix}. \tag{49}$$

The elements of the boundary $\mathfrak{B}$ of the doc $\mathfrak{D}$ are included among all possible sums over subsets of these products. Any such sum may be written as the sum of two terms:

$$d(\mathbf{Y}) = \sum_{y \in \mathbf{Y}_a} D_1(t_1) d_2(y) + \sum_{y \in \mathbf{Y}_b} \overline{D_1(t_1)} d_2(y) \tag{50}$$

and

$$f(\mathbf{Y}) = \sum_{y \in \mathbf{Y}_a} F_1(t_1) f_2(y) + \sum_{y \in \mathbf{Y}_b} \overline{F_1(t_1)} f_2(y) \tag{51}$$

To find the extreme points we need only consider extreme choices for the sums; that is, we need only consider subsets $\mathbf{Y}_{a,b}$ which are the trigger sets for points on the upper boundary $(F(t),D(t))$ for some value of t, the tuning of detector S_2. Since there are two sums involved, there are two tunings, which we may denote as t_{2a} and t_{2b}. Hence the extreme points of the doc for the series case will have the form:

$$D(Y)=D_1(t_1)D_2(t_{2a})+\overline{D_1(t_1)}D_2(t_{2b}) \tag{52}$$

$$F(Y)=F_1(t_1)F_2(t_{2a})+\overline{F_1(t_1)}F_2(t_{2b}). \tag{53}$$

We may see explicitly that the doc of the series case contains that of the two-fold fusion system. In the latter, one of the sets Y_a and Y_b will be either the empty set or the full signal set. For example, when $t_{2b}=0$ the expression reduces to a point in the boundary of

$\mathrm{AND}(\overline{\mathfrak{R}(2)S_1}, \overline{\mathfrak{R}(2)S_2})$.

while when $t_{2a}=1$ it reduces to a point in the boundary of $\mathrm{OR}(\overline{\mathfrak{R}(2)S_1}, \overline{\mathfrak{R}(2)S_2})$.

All the other special cases can be shown to lie within the docs corresponding to either the AND or the OR logic for the fusion case. This confirms that when the second sensor gives up the freedom to let its tuning depend on the signal received from the first sensor, its power is reduced to that of the situation in which each sensor must set its tunings without knowledge of the signal from the other.

4.7 The puzzle of three-fold fusion.

In the case of three detectors we have found that there are 9 non-dominated LOGICS. Three of these are symmetric. The remaining ones form two familes, each of which is closed and transitive under permutations of the three sensors. We can demonstrate, by explicit example that any of the 9 rules may be needed in the general case.

When the three sensors are identical we have not found any cases in which the non-symmetric rules are needed to determine the boundary of the doc. Thus we are led to speculate that perhaps, when all three sensors are the same, only the three symmetric rules are needed to find the extreme points of the doc. However, the argument "symmetric sensors, therefore symmetric solutions" is a dangerous one, as shown in Section 5.1. Thus this question remains open.

5. Specific network results.

5.1 Spontaneous symmetry breaking.

In many cases where the two sensors whose signals are combined at a fusion center are themselves identical, we have found that the optimal tunings of the individual sensors are themselves the same for every tuning of the overall system. This kind of symmetry makes calculations much faster, particularly as one advances to systems with more than two sensors or more than two messages per channel. One may invest considerable energy in the effort to prove that this symmetry holds quite generally, but the efforts are doomed to failure because counter-examples exist.

We exhibit such a counter-example here. The reader will note that the degree of difference between the performance of the system with optimal tunings of the indivudal sensors, and the performance with suboptimal, symmetric tunings, is not large. Thus it may be possible to prove that symmetric tunings represent a heuristic for tuning which comes within some provable discrepancy of the best possible tuning.

The Counter-example.

Consider the sensor defined by the following fundamental table:

$$S = \left| \begin{array}{cccc} d & .375 & .537 & .088 \\ f & .250 & .390 & .360 \end{array} \right| \tag{54}$$

The upper boundary of the doc corresponding to the continuous version of this sensor is given by linear interpolation between the points:

$$\mathcal{S}^{+} = \left| \begin{array}{cccc} 0 & .375 & .912 & 1.0 \\ 0 & .25 & .64 & 1.0 \end{array} \right| \tag{55}$$

By direct calculation one finds that the symmetric tunings for the AND and OR logic at $F = .64$ are in fact both lower than .912, and so surely lower than the best that can be

Figure 7.1 Spontaneous symmetry breaking. The lower curve is the boundary of the doc when both sensors are tuned to the same value. The upper curve is achieved by relaxing that restriction, giving full fusion of the sensors.

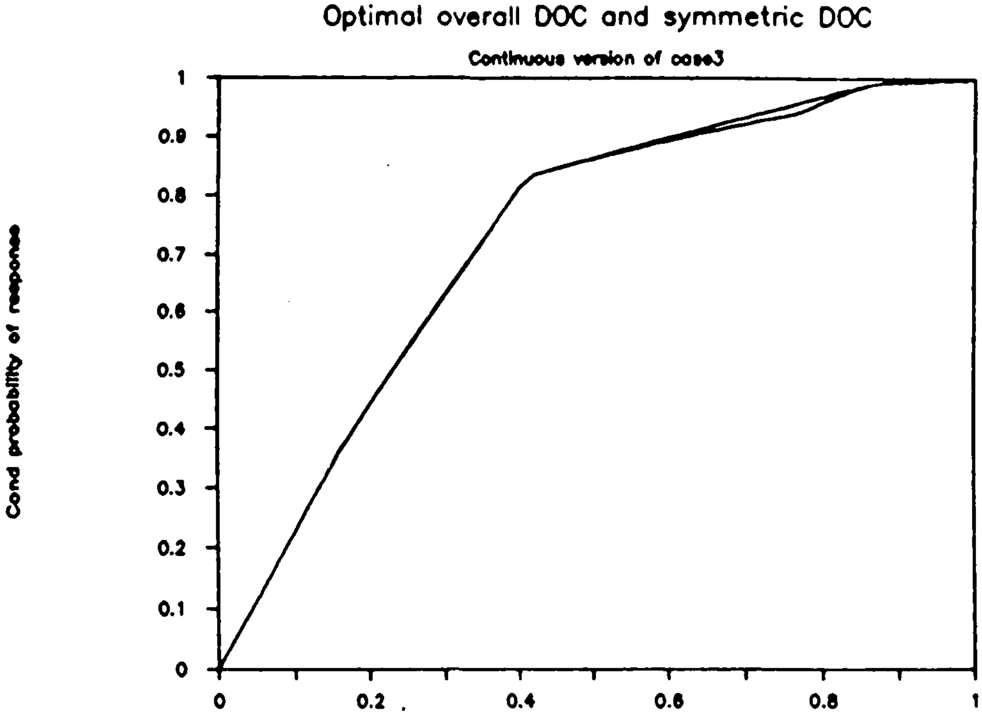
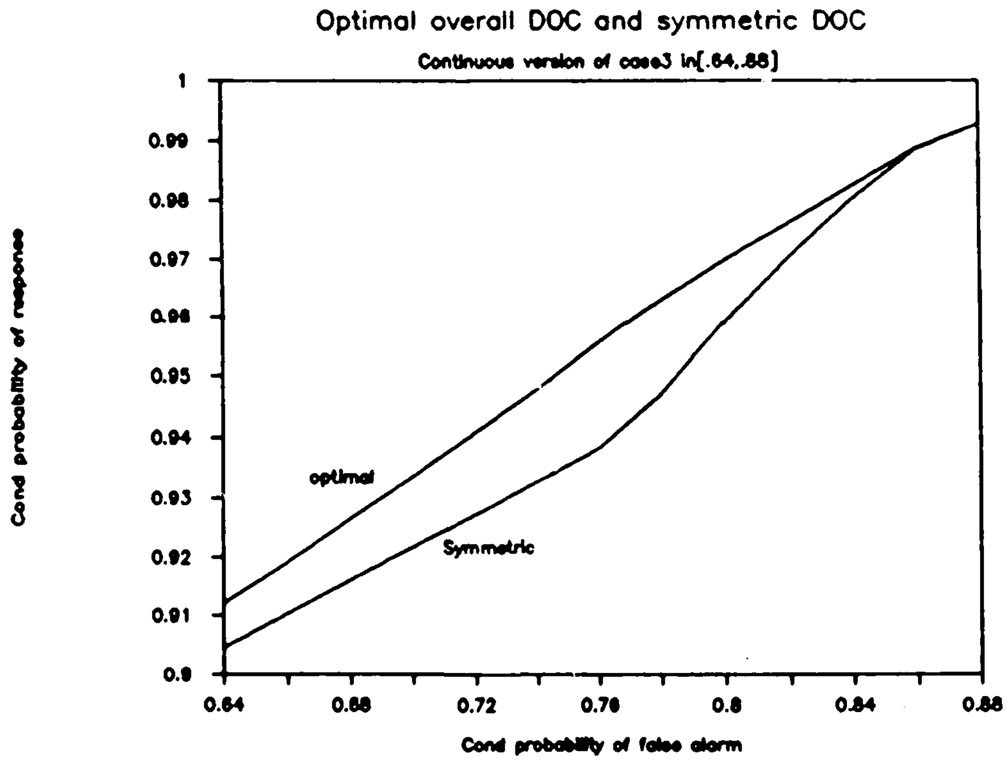


Figure 7.2 Detail of spontaneous symmetry breaking. The largest difference is of the order of 2% in the conditional probability of detection.



achieved with a fusion system. That difference is shown in Figure 7.1,7.2.

5.2 Non-convexity of the doc for fusion systems.

As we shall discuss in detail in Section 6-8, the convexity of the doc $\mathfrak{D}$ for a detector system plays an important role in the solution of decision problems, whether they are characterized in terms of acceptable error rates or by a cost matrix relating actions and hypotheses. Since the solution of any decision problem involves optimization over the doc $\mathfrak{D}$, it is made easier if that region is convex. However, the doc defined by a fusion system is shaped by the fact that the action of fusion corresponds to a discrete sensor, with only 4 points in its doc. For a fusion system, even though the distributed sensors have continuous signal set, and continuous doc, the fusion center itself has a discrete signal set, corresponding to finitely many logics. Examination of the boundary of the doc in example cases shows that the doc is actually the union of several convex sets. Although the intersection of convex sets will also be convex, the union, in general, will not. So it is not surprising to see that there are small regions of non-convexity in the doc of the fusion system. An example is shown in Figure 8.1.

This corresponds to the fusion of two sensors each having the table:

$$S = \begin{vmatrix} d= & .4 & .6 & 0 \\ f= & 0 & .6 & .4 \end{vmatrix} \tag{56}$$

This structure means that each of the sensors can send any of three signals. The first is an unambiguous identification of the desired event; the third is an unambiguous rejection of it, and the middle signal is perfectly ambiguous. As shown in Figure 8.2, there is a substantial dimple in the area where the boundaries of the docs for the two LOGICs cross. The depth of this dimple can be measured in "natural units" corresponding to distance in the Euclidean metric on the doc set. It is approximately 6%. We have calculated the depth of the dimple for all choices of sensor having the general form used here, expressed as a function of the degree of overlap of the response functions. In this case the overlap is 60%, which is close to the location of the maximum. If e denotes the area of overlap, the depth of the dimple may be shown, by

Figure 8. Non-convexity of the fusion doc. The individual sensors each have doc boundaries corresponding to the curve labeled DOC. The fusion of two sensors with binary messages has the boundary labelled MAX DOC. There is a clear "dimple" or non-convex region at the point where the LOGIC changes.

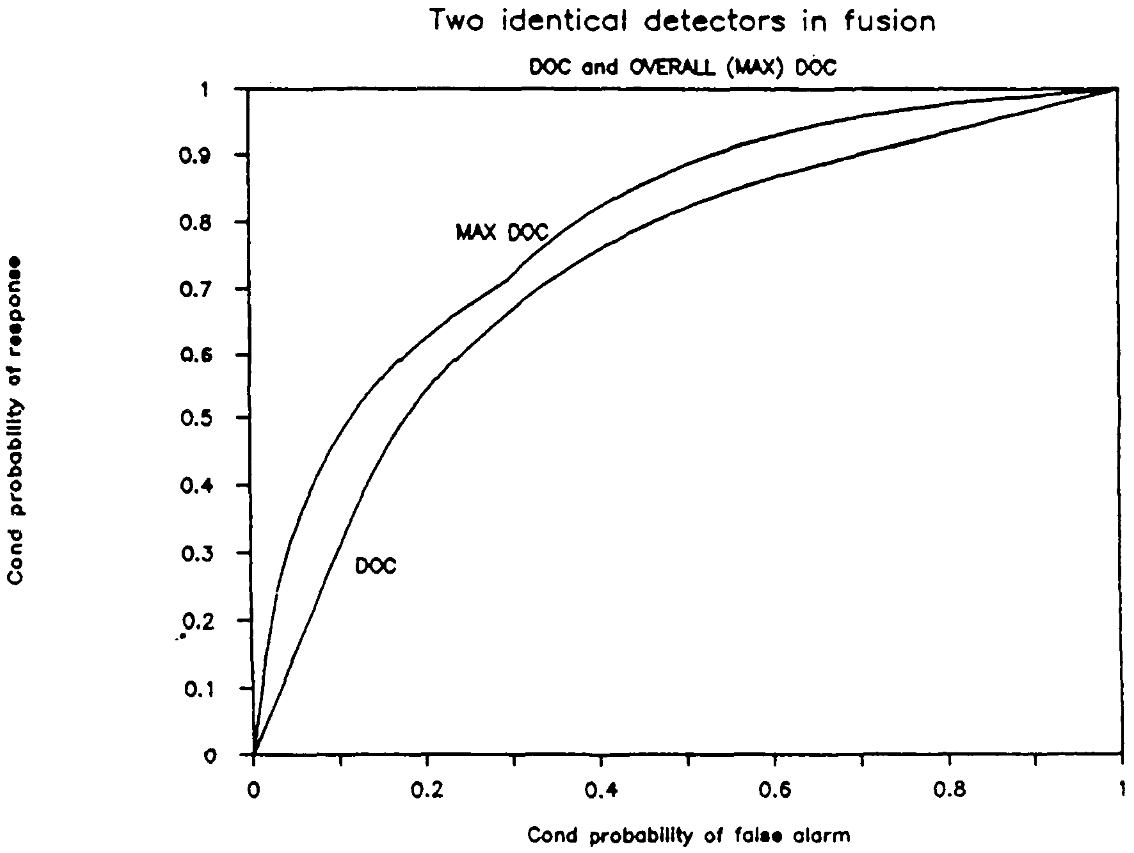
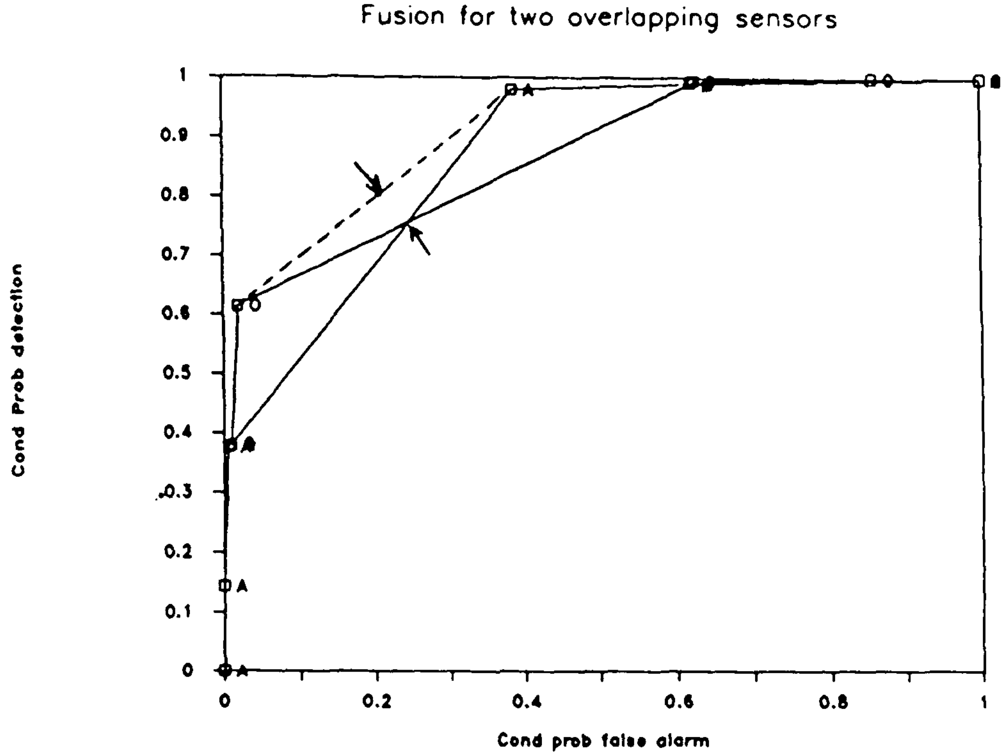
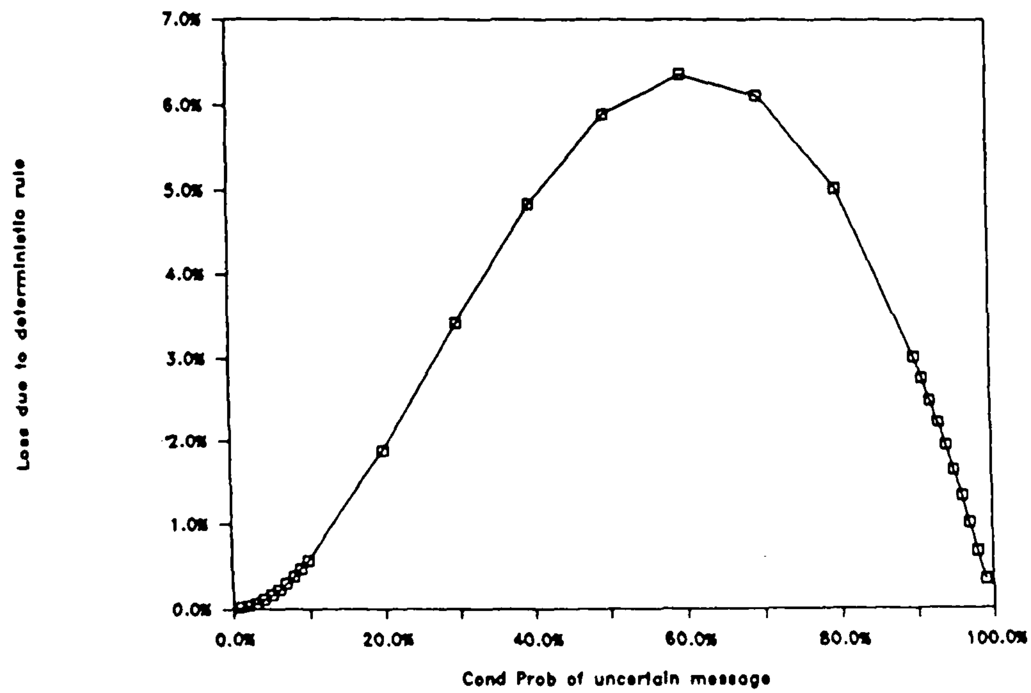


Figure 8.2 Nonconvexity of the doc in fusion. A fairly extreme case occurs when there is a 60% chance that each of the sensors will give a completely ambiguous signal y∈Y. The dimple is approximately 6%.
Figure 8.3 Nonconvexity as a function of ambiguity. The depth of the dimple, measured as Euclidean distance in the doc space, is given as a function of the conditional probability of the ambiguous signal from the component detectors.





elementary geometry, to be:

$$\mathrm{GAP}=\frac{\epsilon^2}{\sqrt{2}}\,\frac{1-\epsilon}{1+\epsilon} \tag{57}$$

which achieves its maximum at:

$$\epsilon=(\sqrt{5}-1)/2=0.618... \tag{58}$$

One expects that non-convexity will be the rule for fusion systems, unless a single logic dominates all of the others, and this example of a typical model of sensor imperfection suggests that the effects may be substantial. (See also Figures 8.2 and 8.3)

5.3 The series topology.

In general, the simplest series structure is represented by $\overline{\mathfrak{R}(2)S_1}\otimes S_2$. One sensor sends a binary signal to another. One natural question is whether, if one sensor is definitely better than the other, the good sensor should be placed "upstream" or "downstream." As we discuss in Section 6, a sensor S is definitely better than another S' if and only if the doc $\mathfrak{D}(S)$ includes the $\mathfrak{D}(S')$. We illustrate some of the complexity of this problem with a simple finite example. Let two sensors be described by the tables:

$$G=\begin{bmatrix} .40 & .35 & .15 & .10 \\ .25 & .25 & .25 & .25 \end{bmatrix} \tag{59}$$

and:

$$B=\begin{bmatrix} .40 & .30 & .20 & .10 \\ .25 & .25 & .25 & .25 \end{bmatrix}. \tag{60}$$

It is readily verified that G is better than B in that sense.

There are three non-trivial reductions which may be applied in this case — combination

of the first two columns, the first three columns, and the last three columns. We may then form the direct sensor product of the reductions of G with B, and vice versa. This yields a set of points which are the extreme points of the composite structure $\overline{\mathfrak{R}(2)U} \otimes D$. We use "U" to represent the "upstream detector" and "D" to represent the "downstream detector." The results are most easily seen in the graphs of Figures 9.1, 9.2, 9.3 which show the complete upper boundaries $\mathfrak{B}^{+}$ of the two composite docs. These have been verified by direct calculation based on the piecewise linear formulation corresponding to the discrete sensors G and B. The enlarged views show that in the neighboorhood of $F=0.5$ it is "better" to have the better sensor upstream, while in the neighborhood of $F=0.625$ it is better to have the poorer sensor upstream.

As with the case of the spontaneous symmetry breaking the effects are small, and we do not know whether they can be shown to be small, in this sense, in every case. Of course it must be remembered that differences on the doc graph are multiplied by some scale factor depending on the importance of the problem to which the sensor system is applied.

Figure 9.1 Two sensors in series. The individual sensors have the upper boundaries 1, which is better, and 2, which is poorer. The two possible choices for which sensor is upstream yield two composite docs, whose upper boundâries are close to each other.
Figure 9.2 Detail of Figure 9.1. In the vicinity of 50% false alarm rate, the performance of the system with the better detector upstream is superior.

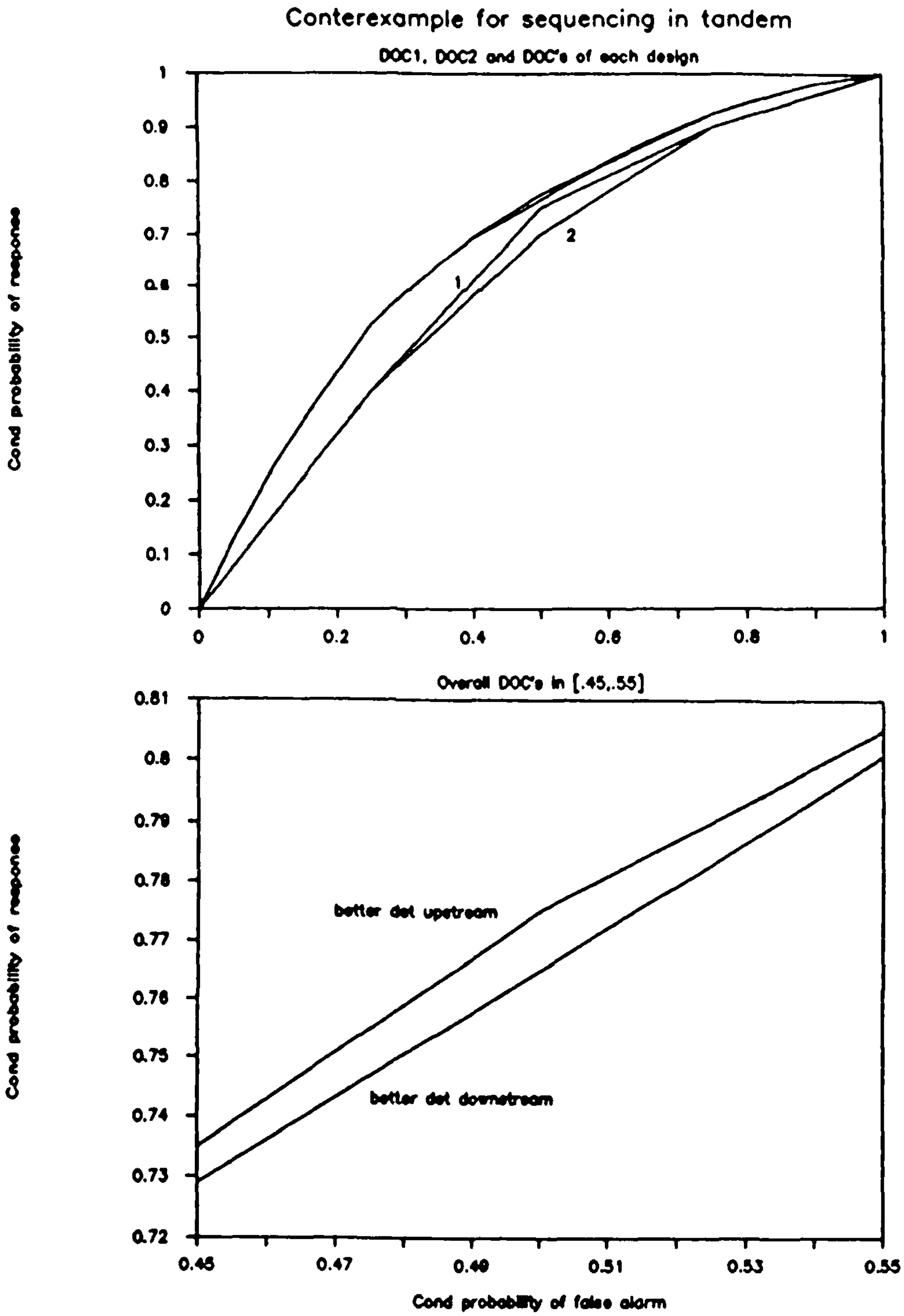
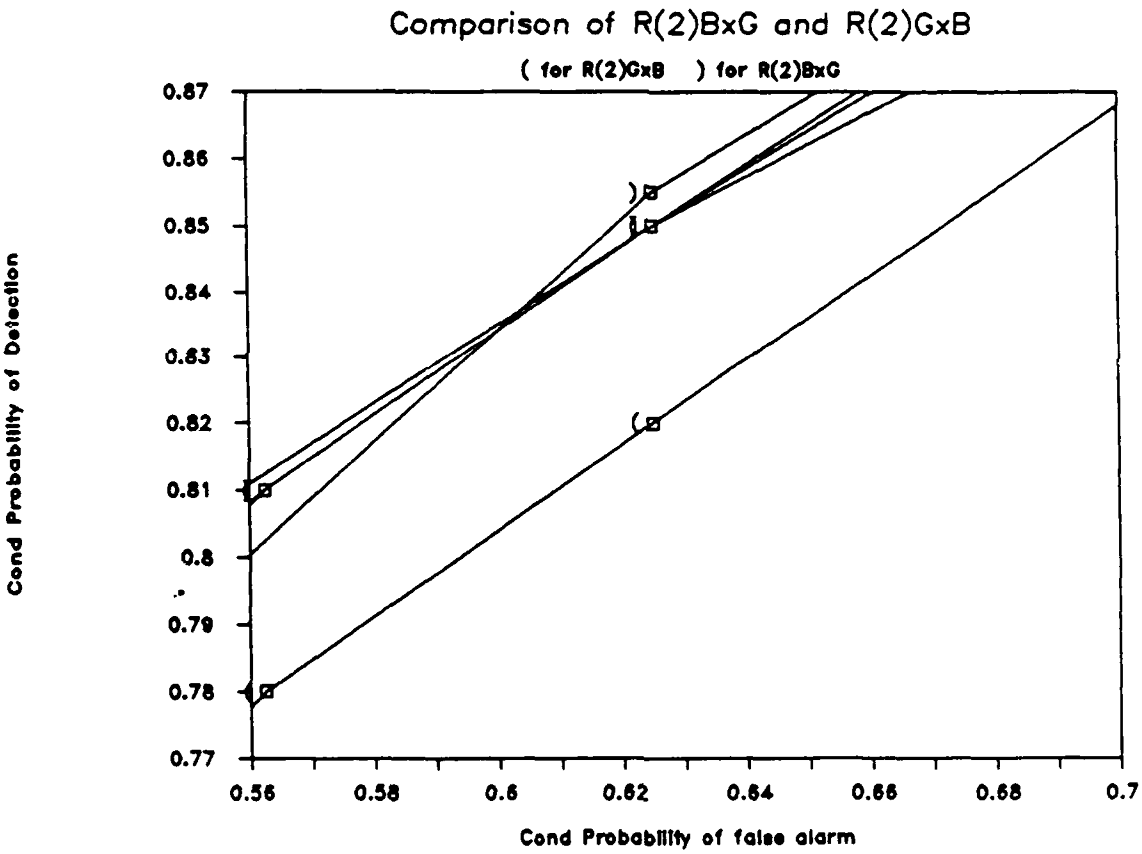


Figure 9.3 Detail of Figure 9.3. In the vicinity of 62.5% false alarm rate, the performance of the system with the poorer detector upstream is superior. Possible tunings of the system with the poorer detector upstream are labeled ")". There are six variant tunings of the two possible configurations, but several of them overlap in this small region.



5.4 General procedures for the calculation of any network.

The ideas set forth in Sections 1-3 permit us to provide an algorithm for the discussion of any network in which signals flow only one way, and there are no closed loops. We have seen that such a network can be represented by some combination of the basic operations of the sensor calculus, restriction $\mathfrak{R}(k)$ and full product $\otimes$. We confine ourselves to the case of two hypotheses and two-fold signals.

The doc boundaries $\mathfrak{B}$ of the root sensors may be given in either analytical form or numerical form. Thereafter all calculations will produce numerical form. The result of any such calculation can be thought of as a table enumerating the points of the boundary $\mathfrak{B}$, and the tunings of the component detectors that give rise to them. For example, when the operation is a binary fusion $\overline{\mathfrak{R}(2,t_1)S_1} \otimes \overline{\mathfrak{R}(2,t_2)S_2}$ the table contains entries:

$$F \qquad D \qquad t_1 \qquad t_2 \quad \mathsf{LOGIC} \tag{61}$$

where t_1 and t_2 and LOGIC are the coordinates of the optimal solutions to the problem:

$$D = \max_{\mathsf{LOGIC}} \quad \max_{t_1, t_2} Q_{\mathsf{LOGIC}}\Big(D_1(F_1(t_1)),\, D_2(F_2(t_2))\Big) \tag{62}$$

subject to the conditions:

$$Q_{\mathsf{LOGIC}}\Big(F_1(t_1),\, F_2(t_2)\Big) = F \tag{63}$$

and:

$$\Big(F_1(t_1),\, D_1(t_1)\Big) \epsilon \mathfrak{B}(S_1) \tag{64}$$

$$\Big(F_2(t_2),\, D_2(t_2)\Big) \epsilon \mathfrak{B}(S_2). \tag{65}$$

In practice, for the case of two hypotheses and two actions, the value of $F_i(t_i)$ may be used to represent the tuning as well. If, for reasons of economy, or reliability, the choice of LOGIC is fixed then the table will only contain the values of (F,D,t_1,t_2).

For the case of fusion of more than two sensors, the general form will be the same, but the enumeration of LOGICs will be more complex.

For the series structure $\overline{\mathfrak{R}(2,t_u)U} \otimes D$ the table is slightly more complicated,

containing:

$$F \quad D \quad t_u \quad t_{d1} \quad t_{d2} \tag{66}$$

where t_u, t_{d1} and t_{d2} solve the problem:

$$D = \max_{t_u,\ t_{d1},\ t_{d2}} D_U(t_u)D_D(t_{d1}) + [1 - D_U(t_u)]\ D_D(t_{d2}) \tag{67}$$

subject to the conditions

$$F_U(t_u)F_D(t_{d1}) + [1 - F_U(t_u)]\ F_D(t_{d2}) \leq F \tag{68}$$

and:

$$\left(F_U(t_U),\ D_U(t_U)\right) \epsilon\, \mathfrak{B}(U) \tag{69}$$

$$\left(F_D(t_{d1}),\ D_D(t_{d1})\right) \epsilon\, \mathfrak{B}(D) \tag{70}$$

$$\left(F_D(t_{d2}),\ D_D(t_{d2})\right) \epsilon\, \mathfrak{B}(D) \tag{71}$$

Such tables may be used to determine the doc boundary $\mathfrak{B}$ for any composite into which these composites enter as components. In the practical application of such a system all the tables must be maintained available for use. The overall detector system boundary $\mathfrak{B}$ is used to determine the overall optimal tuning, as described in Sections 6-8. That tuning is then looked up in the overall system tuning table to determine the optimal tunings of the major subcomponents. This lookup process is iterated until the tuning of each fundamental sensor in the network has been determined, as well as the tunings of the intermediate sensors.

6. The link between decision making and the doc.

6.1. Bayesian Formulation.

In the Bayesian formulation a problem is described in terms of probabilities, actions and hypotheses. In general there is a cost matrix $C(a,h)$ defined for $a\epsilon\mathbf{A}$, the set of actions and $h\epsilon\mathbf{H}$, the set of hypotheses. In general one might permit the cost function to also depend on the probability that each of the several hypotheses is true, in which case we might write it as $C(a,p)$, where $p=(p(h=1),\ldots,p(h=H))$. The probabilities have some specified values p^0 prior to the observation, and are updated to the values p, by the observations of the sensor network. In the general case the cost may be a non-linear function of p. For example, the cost may increase more rapidly as the leakage through a defensive system increases, and the number of survivors decreases. Thus improvement in detection yields diminishing returns. In another setting, improvement in detection will, in general, bring diminishing returns because the precision of a measurement increases only as the square root of the number of detections.

We restrict ourselves in this paper to the customary engineering cost model, in which the cost function is linear in the individual probabilities, and may be written as:

$$C(a,p)=\sum_{h\epsilon\mathbf{H}} C(a,h)p(h). \tag{72}$$

In this case, the cost of choosing a given action may be expressed as:

$$EC(a\epsilon\mathbf{A})=\sum_{h\epsilon\mathbf{H}} Pr(a \text{ and } h)C(a,h). \tag{73}$$

If the action a is taken whenever the signal y is in the trigger region $\mathbf{Y}(a)$, this becomes:

$$EC(a)=\sum_{h\epsilon\mathbf{H}} p_h^0\, f_h(\mathbf{Y}(a))C(a,h). \tag{74}$$

Note further that, for this formulation, in which the expected cost is to be minimized without additional constraints, the cost matrix $C(a,h)$ and the prior probability enter only in a

single combination: $W(a,h)=p_h^0 C(a,h)$. This can be thought of as a vector, labelled by a, whose several components are indexed by h. With this perspective, the expected cost is simply the inner product of the vector $f(\mathbf{Y}(a))$, whose components are $f_h(\mathbf{Y}(a))$, and the vector $W(a)$.

Thus solution of the Bayesian decision problem amounts to finding, for each a, the tuning $\mathbf{Y}(a)$ which minimizes the overall expected cost. In the general case this will be accomplished by assigning each element $y \epsilon \mathbf{Y}$ to that action $a(y)$ for which $W(a(y)) \leq W(a')$ for all other actions a'. We concentrate on the case of exactly two hypotheses and two actions. In this case we need only specify one trigger set, $\mathbf{Y}(a=1)$, with the other, $\mathbf{Y}(a=0)$, defined by complementation. Similarly, the space in which the vectors f and W lie has only two dimensions. As in Section 4, we will refer to its axes as the f and d axes, corresponding to $h=0$ and $h=1$ respectively. The components of f are (f,d).

It can be shown by direct calculation [Blankenbecler,Kantor88] that <u>for this case</u> the expected cost depends only on certain differences, which may be thought of as the cost of two kinds of error:

$$C(h=1)=C(a=0,h=1)-C(1,1) \tag{75}$$

and

$$C(h=0)=C(a=1,h=0)-C(0,0). \tag{76}$$

We assume, without loss of generality, that both of these quantities are positive. (If they are both negative we should relabel the actions. If they have opposite signs then one of the actions is to be preferred whatever the state of nature, and no sensor system is needed.) This particular simplification is unique to the case of two actions, in which the difference vector $R=\mathbf{W}(\mathbf{a}=1)-\mathbf{W}(\mathbf{a}=0)$ can be used to determine whether $W(a=1)\cdot f(y) \lesssim W(a=0)\cdot f(y)$.

By direct calculation we find that the expected cost may be rewritten as:

$$EC=\sum_{h=1}^{H}\sum_{a=1}^{0} p_h^0 f_h(\mathbf{Y}(a))C(a,h) \tag{77}$$

$$=\sum_{h=1}^{H}\left[p_h^0 f_h(\mathbf{Y}(1))C(1,h) + p_h^0 f_h(\mathbf{Y}(0))C(02,h)\right] \tag{78}$$

$$= \sum_{a=1}^{0} W(a)\cdot f(\mathbf{Y}(a)) \tag{79}$$

We use the fact that there are only two possible actions to write $f(\mathbf{Y}(2))=E-f(\mathbf{Y}(1))$, leading to

$$EC= W(0)\cdot E+(W(1)-W(0))\cdot f(\mathbf{Y}(1)). \qquad E=(1,1) \quad (H\ terms). \tag{80}$$

Thus the problem of minimizing the expected cost is the same as minimizing the value of the second dot product. This can be visualized as sweeping a hyperplane perpendicular to $\mathbf{R}=\mathbf{W}(1)-\mathbf{W}(0)$ across the doc until it reaches an extreme point. When the dot product is as small as possible the corresponding choice of the tuning $\mathbf{Y}(a=1)$ is optimal. We show this construction graphically. Note that the two terms of R are, using "f" for "0", and recalling that a=0 means "do not act":

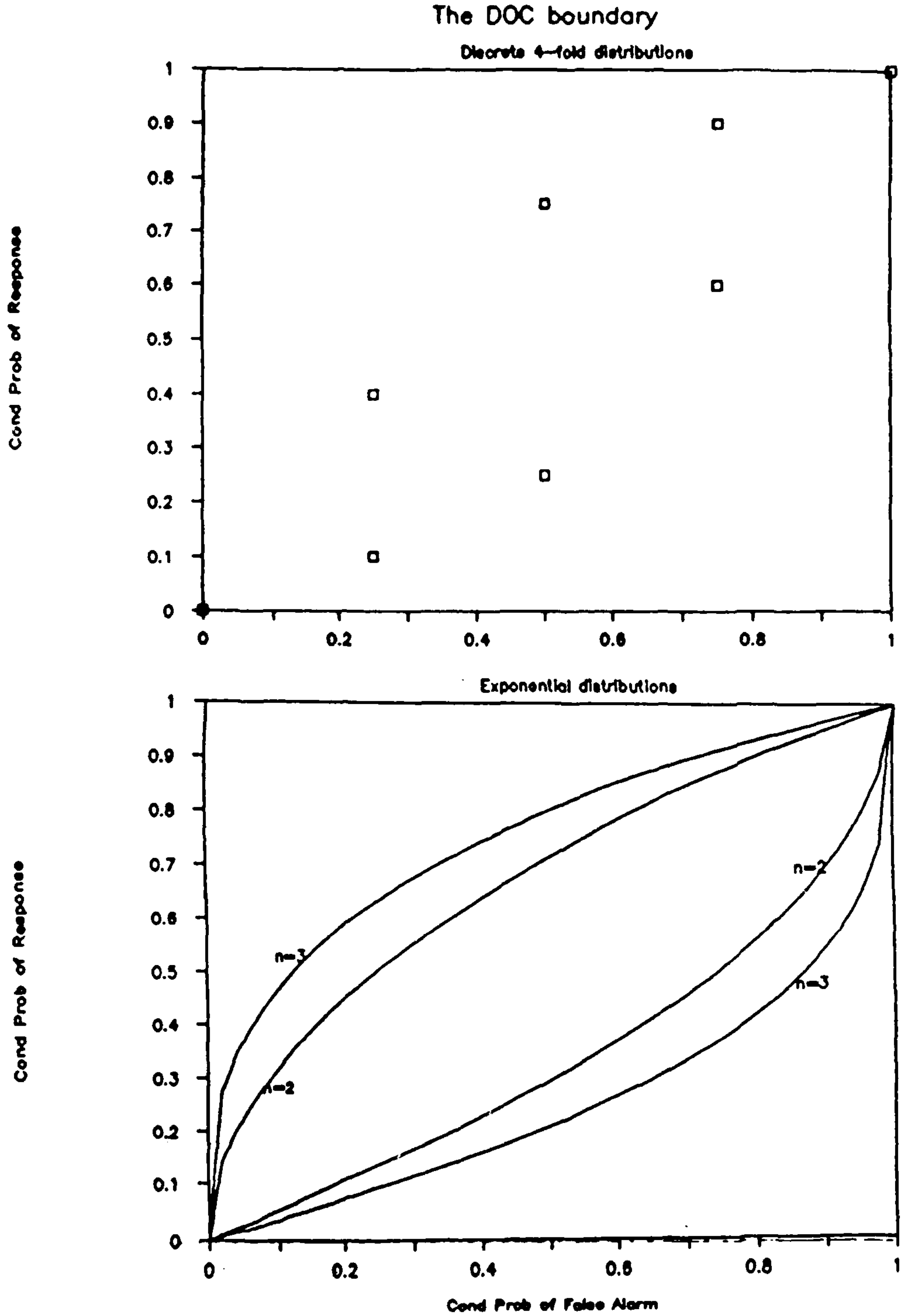
$$R_d=p_1^0 C(1,1)-p_1^0 C(a=0,1) \quad h=1 \text{ (“}d\text{”)} \tag{81}$$

$$R_f=p_0^0 C(1,2)-p_0^0 C(a=0,2) \quad h=2 \text{ (“}f\text{”)} \tag{82}$$

These represent the *a priori* risk associated with the two possible states of nature. Presumably the first is negative and the second is positive. Their ratio determines the slope of the line that sweeps across the doc. For $|R_d| \ll R_f$ the line is nearly vertical, which favors a tuning very close to (0,0). This is reasonable since the risk of an incorrect response (false alarm) is relatively great, inhibiting us from action. Conversely, when $|R_d| \gg R_f$ tunings close to (1,1) are preferred because a miss would be very costly.

No matter what the nature of the doc — be it discrete or continuous — the solution to Bayesian problems will be found at the extreme points of the doc. If it is discrete these are isolated points, as in Figure 10.1. If it is continuous all the points of the boundary are extreme points (Figure 10.2). Thus, in one way or another, everything that we need to know about the doc is contained in a "listing" of its extreme points. This may be given in closed analytic form

Figure 10.1 Bayesian problems: Discrete case. The cost and prior probabilities determine a direction, represented here by level lines. The solution to the problem is to move as far in that direction as possible, without leaving the doc. As the direction rotates, the optimal tuning remains "stuck" at a vertex of the doc, until it is ready to jump to another one.
Figure 10.2 Bayesian problems: Continuous case. When the boundary of the doc is continuous, the level line for the optimal tuning rolls around the doc, and the tuning point changes continuously.



(rarely), in a tabular form with interpolation rules, or by direct enumeration.

To sum up, the solution of any Bayesian problem is reduced to complete knowledge of the boundary of the doc. Since the boundary of the doc also provides all the information needed to carry out the operations of the sensor calculus, we consider an alternative way to characterize it.

6.2 The Neyman Pearson Formulation.

In general, the extreme points of the doc are all those points through which a hyperplane may be passed without including any interior points of the doc. In the general case the interior points are all convex combinations of extreme points that are not themselves extreme points. One way to enumerate the extreme points is to consider all possible Bayesian problems, and find the solutions for each. Because the information about prior probabilities and about costs enters only through a single ratio, this would be a highly redundant enumeration. A more efficient approach is to consider all values of the determining ratio R_d/R_f.

Yet another method is to trace out the extreme points by gradual relaxation of an artificial constraint. This approach is familiar from statistics, where it defines the operating characteristic of a test, and has given rise to the name "Receiver Operating Characteristic." Interestingly enough, this terminology has made its way back into statistics as well. [Kraemer88, Swets72, Swets88]. The general theory of most powerful tests was developed by Neyman and Pearson [Neyman42] and so we refer to this approach as the Neyman-Pearson formulation of the optimal discrimination and decision problem.

Points on the boundary of the doc can be characterized as either:

$$D(F) = \max_{(f,d)\in doc,\ f\leq F} d \tag{83}$$

or:

$$F(D) = \min_{(f,d)\in doc,\ d\geq D} f. \tag{84}$$

This approach has been used to provide independent proofs of the convexity of the doc, and of the fact that the boundary is piecewise differentiable. [Cherikh88]. When the doc is continuous, the function $D(F)$ has a simple relation to the effective cost vectors W at the optimum tuning. Since the tuning point t^* is on the boundary we have:

$$(W(a=1)-W(a=2))\cdot(F(t^*),D(t^*)) \tag{85}$$

is a minimum, or:

$$\frac{d}{dt^*}\left[F(t^*)R_f - D(t^*)\left|R_d\right|\right]=0. \tag{86}$$

or:

$$F'(t^*)R_f = D'(t^*)\left|R_d\right|. \tag{87}$$

But $D'(t^*)/F'(t^*) = dD/dF\big|_{t^*}$. (88)

That is, the slope of the boundary $D(F)$ at the tuning point, is given by $R_f/\left|R_d\right|$. In practice $D'(F)$ is often most easily found in the parametric form. Because the slope is itself monotonically decreasing, many fast algorithms exist for finding the optimal tuning.

6.3 Constrained Optimization.

The formulation just given, for the Neyman Pearson problem, represents the simplest kind of constrained optimization. There are realistic situations in which another kind of constraint arises. Consider a situation in which there is an expected series of incidents, numbering I in all. Suppose that the available budget of responses is B, and that the prior probability is that a fraction p_1^0 will be "true events" corresponding to $h=1$, while a fraction p_2^0 will be "non events" corresponding to $h=2$. Then when the system is tuned to the operating point $\left(F(t^*),\ D(t^*)\right)$ the expected total number of responses will be:

$$ER = p_1^0 D(t^*) + p_2^0 F(t^*). \tag{89}$$

If ER is less than B there is no problem. On the other hand, if the number of responses exceeds the budget allowed then some fraction of the incidents will not, in fact, be responded to. The effective performance in this case is reduced to $(B/R)(F(t^*),D(t^*))$. This point is a

convex combination of the points $(F(t^*),D(t^*))$ and $(0,0)$ and so is interior to the doc. It is therefore not optimal for any choice of the prior probabilities and cost information. Thus the optimal tuning will be the tuning for which the expected number of responses is equal to the budget.

So, generally, the solution to the problem:

$$\max \mathbf{W}\cdot f \tag{90}$$

subject to the constraints:

$$f \in \text{doc} \tag{91}$$

$$\text{and} \quad p^0 \cdot f \leq B/I \tag{92}$$

will lead to a point which is on the boundary of the doc and, for some choices of the cost matrix, also on the line given by $p^0 \cdot f = B/I$.

If the doc is continuous this does not pose any problems. Every point on the boundary of the doc is accessible by a suitable choice of the tuning $\mathbf{Y}(a)$. However, if the doc is discrete it may happen that the intersection of the boundary (strictly speaking, of the convex hull of the extreme points) with the constraint given by Equation (92) will not be a point of the doc. In the most general analyses of optimal design of experiment [Blackwell54] this is dealt with by using a "mixed strategy." Under a mixed strategy, points on the line connecting two elements of the doc are achieved by using each of the corresponding strategies a fixed fraction of the time, with random selection of which strategy is to be used at any given time.

For a single sensor one may implement such a random strategy by broadening the bins into a continuum, and then choosing a tuning which effectively mixes the bins in fixed proportions. For example, the sensor described by the table:

$$S_1 = \begin{bmatrix} 1 & 2 \\ .75 & .25 \\ .25 & .75 \end{bmatrix} \tag{93}$$

may be replaced by a sensor with continuous signal set $\mathbf{Y}=[0,1]$ and the response functions:

$$f_0(y)=1.$$
$$f_1(y)=\begin{cases} 3 & 0\leq y\leq .25 \\ 1/3 & .25\leq y\leq 1 \end{cases} \tag{94}$$

The graph of the upper boundary of the doc is:

$$D(F)=\begin{cases} 3F & 0\leq F\leq .25 \\ .75+(F-.25)/3 & .25\leq F\leq 1. \end{cases} \tag{95}$$

It is clear that this doc has no gaps in its boundary. In terms of the original sensor table, a tuning such as $F=.5$ corresponds to a mixture:

$$(1/3)S_1(\text{tuned to } \mathbf{Y}(a)=\{1,2\})+(2/3)S_1(\text{tuned to } \mathbf{Y}(a)=\{1\}) \tag{96}$$

with the false alarm rate:

$$F=(1/3)1+(2/3)(.25)=0.5 \tag{97}$$

and the detection rate:

$$D=(1/3)1+(2/3)(.75)=5/6. \tag{98}$$

7. **Continuity and discontinuity in the behavior of network detector systems.**

It has been noted elsewhere [Blankenbecler,Kantor88] that even for a simple model problem, it may happen that the tuning of a fusion system jumps discontinuously when the LOGIC changes from AND to OR. At the same time, it was observed that the optimal cost corresponding to the best tuning does not exhibit a discontinuity. We are now in a position to explain both of these phenomena, and to comment on their significance for the optimal design of distributed systems.

First we note that, for a discrete sensor, there will be discontinuities of tuning, as the line representing the constant value of the cost "rolls around" the boundary of the doc, touching at the extreme points. However, the cost associated with the best tuning for a given value of R is continuous as the direction of R varies. This is because, when R is such that either of two tunings is optimal (i.e. the line of constant $R \cdot f$ is an extreme edge of the doc) then the cost associated with each of the the two tunings is the same.

Exactly the same phenomenon can occur when the doc $\mathfrak{D}$ of a fusion system exhibits non-convexity which, as we discussed in Section 5, is a general occurrence. We see that for a certain critical value of the vectors **W** the extreme value of the cost will occur at two distinct points on the boundary of the doc, corresponding to two different choices of the logic. With simple fusion, as the difference vector **R** moves the optimal tuning will jump suddenly from the value appropriate for the AND logic to that appropriate for the OR logic. The "cost" at the minimum will not show any discontinuity because the distance between two parallel lines, measured in any direction, is the same no matter where, along the parallel lines, the measurement is taken.

In practice, this could have very serious consequences. A network will, in general, be tuned to our best present understanding of the costs and prior proabilities. If the general characteristics of the sensors are such that the optimum is at or near this point of discontinuity, then we are less confident than we would otherwise be that our tuning is the best one. To take the worst case: suppose that we had to tune right at the ambigous point. We must make a choice of LOGIC, and of the tuning of the individual sensors. If we select, for

example, AND, and the actual situations (costs, priors) is such that the tangent line is a little steeper than we think, everything is fine. We will be operating slightly above the optimium tuning (F(t*), D(t*)), but not very much. However, if the tangent is slightly flatter, we will be operating quite far from the optimal point, with corresponding loss in system performance. As the slope of the tangent goes to zero the cost of wrong LOGIC falls to zero. The cost due to being on the wrong lobe (that is, choosing the wrong LOGIC) is not the same as the difference between the two functions $D_{AND}(F)$ and $D_{OR}(F)$ measured at the same point F.

An example is given in Figures 11.1, 11.2, based upon the sensors described in the imbedding model of [Blankenbecler, Kantor88]. The fundamental table is:

$$\begin{bmatrix} Y=[0,\infty] \\ f_1(y)=\frac{n+1}{n+1-z}e^{-y}(1-ze^{ny}) \\ f_0(y)=ne^{-ny} \end{bmatrix} \tag{99}$$

The upper boundary $\mathfrak{B}^+$ may be given in closed form:

$$D(F)=\frac{n+1}{n+1-z}F^{1/n}\left(1-\frac{zF}{n+1}\right) \tag{100}$$

In plotting the dependence of cost on the slope of R it is convenient to use a reduced measure of cost: $J\equiv f-|R_d/R_f|d$. The true value of the difference in performance depends upon the scale factor R_f, which may be extremely large. In figures 11.1 and 11.2 the independent variable is the angle between the line of constant cost and the vertical axis. This angle $arctan(|R_d/R_f|)$ varies from 0 to $\pi/2$ as the optimal point sweeps around the upper boundary of the doc.

Figure 11.1 Optimal tuning in fusion. For the embedding model of Equations (99,100) the optimal tuning is always symmetric. For two values of the parameters n and z we show that as the direction of the cost minimization rotates, the tuning changes discontinuously, as the logic changes from AND to OR. The tuning point jumps from one lobe of the boundary of the doc to another.

Figure 11.2 Expected cost in fusion. For the same cases as in Figure 11.1, the expected cost, measured in reduced units, does not show any discontinuity.

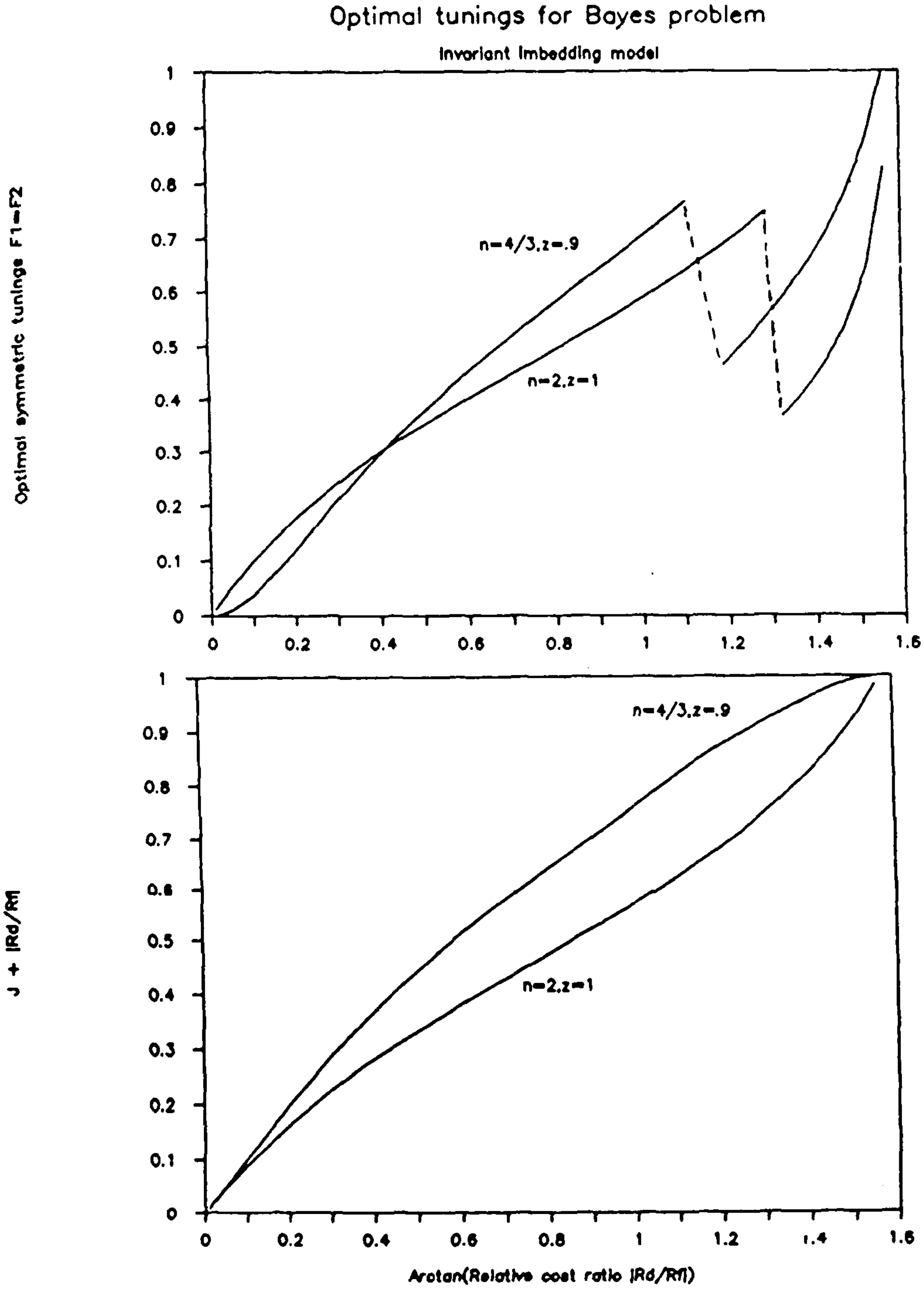
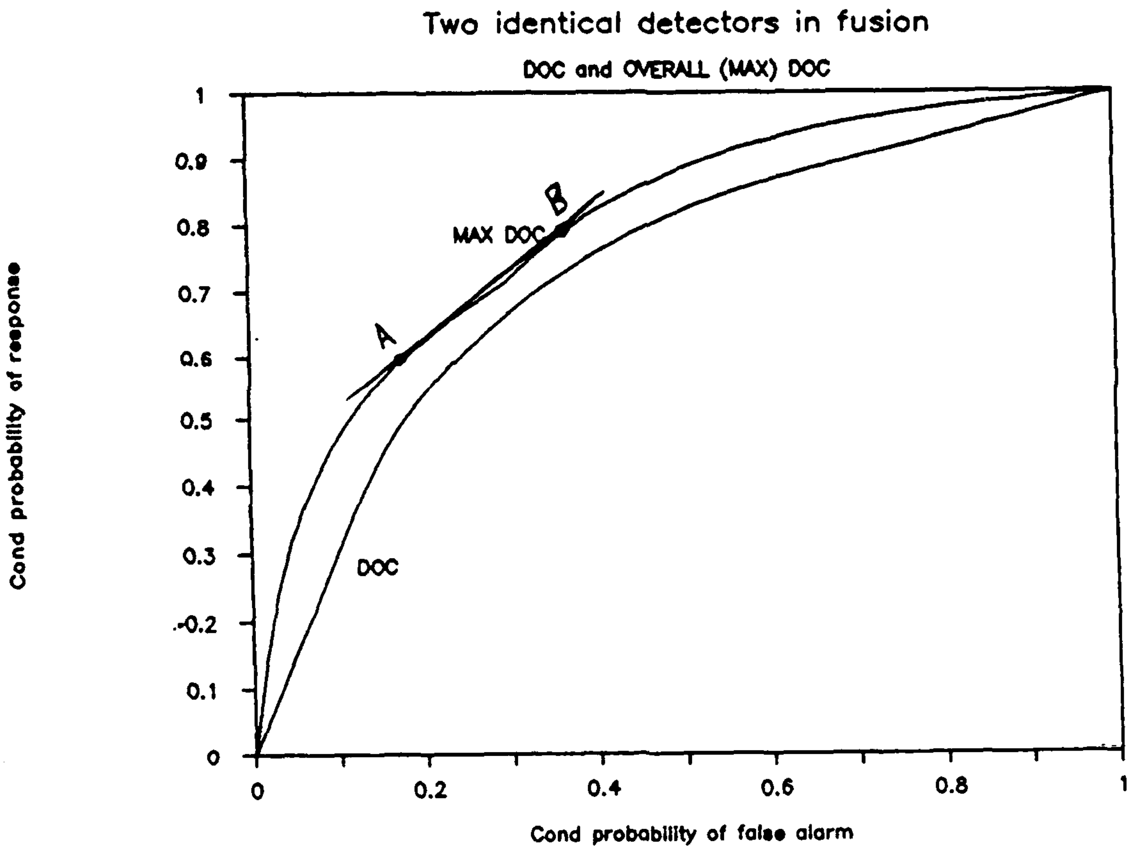


Figure 11.3 Tuning and cost in fusion. The critical point corresponds to the line AB. When the critical value is reached the tuning jumps from point A to point B, while the reduced cost, J, does not jump.



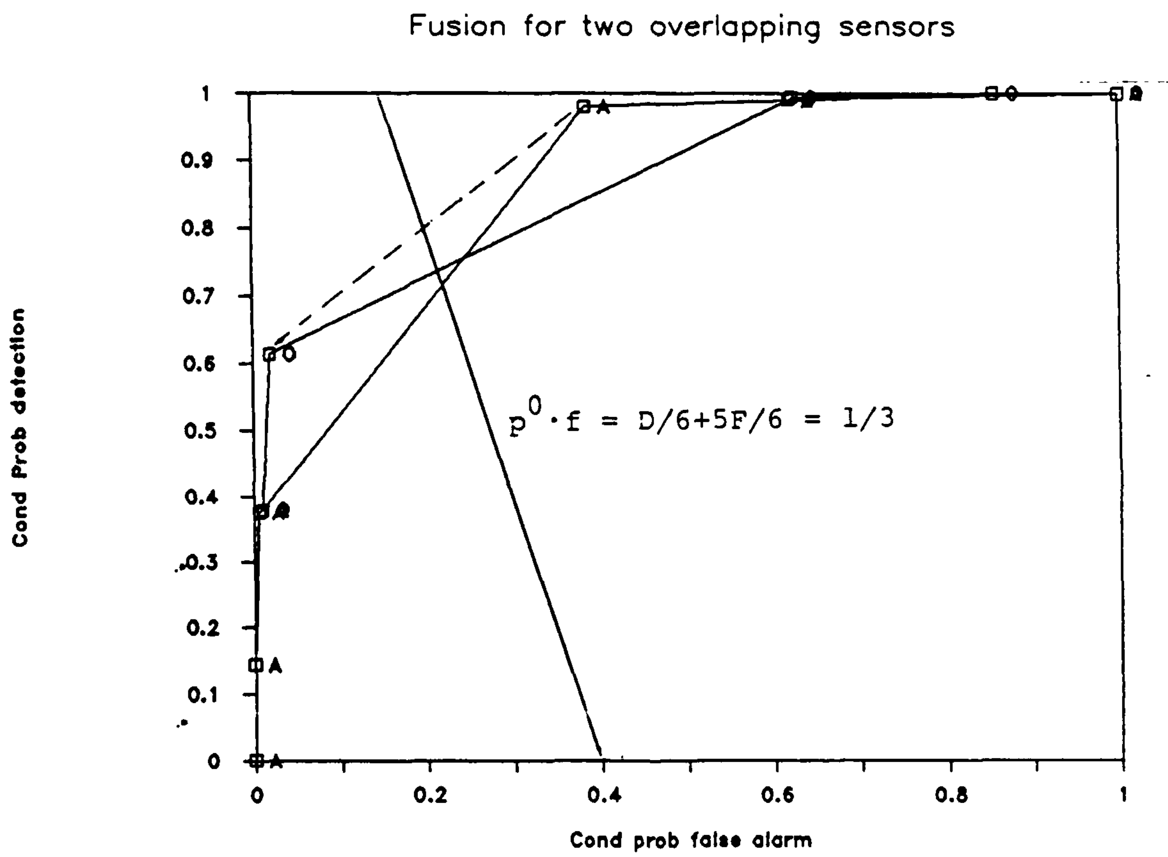
8. Resource constraints and mixed strategies.

We have seen in Section 7 that the finiteness of the set of LOGICs leads fusion systems to exhibit some of the discontinuities of discrete systems. As might be expected, this problem also affects the situation of constrained resources. We saw in Section 6 that, when resources are constrained, the optimal tuning may correspond to a point which is not in the doc of a discrete system. It can, in some cases, be achieved by a mixed tuning, or by a suitable broadening of the signal set, which amounts to the same thing.

In the case of a fusion system, the problem manifests itself as shown in Figure 12. The resource constraint passes through the "dimple" in the boundary of the doc. With mixed strategy one could achieve the value corresponding to the point Q. With a pure strategy (definite tuning) one cannot do better than the point P at which the two boundaries (corresponding to the LOGICs AND and OR) meet. In dimensionless units, the added cost we must bear is given by the depth of the dimple. The maximum perpendicular distance from the boundary of the doc to the line segment forming the convex hull across the dimple is an upper bound (in these absolute units) for the added loss due to the non-convexity of the full doc.

In response to this problem one might ask why we do not propose that a mixed strategy be used, to avoid the added cost. The problem, it seems to us, is that mixed strategy requires coordinated random retuning at each of the distributed sensors, as well as at the fusion center. The communication costs of coordinating the retuning are likely to be higher than the cost of adding to the communication capacity of the system itself, with corresponding gain in system performance. Thus, in the design stages, one should avoid constructing fusion systems for which the dimples in the overall system doc are likely to be in regions of operating interest.

Figure 12. Resource constraints and fusion. Using an earlier example we show that the resource constraint line may pass through the dimple of the doc for a fusion system, resulting in sub-optimal performance.



9. The problem of team action

In the preceeding sections we have concentrated on the case of two actions and two hypotheses. Another problem of some interest [Tenney81] is that in which there are only two actions to be taken, but each of several agents may take them independently. This problem may be discussed, with considerable complication, by using the language of updated probabilities and likelihood thresholds. However, the same constructs that we have used above also make it easy to discuss this problem.

To fix notation for this section, let F represent the probability of false alarm and $D(F)$ represent the probaility of detection at a particular sensor. (In other words, we let F itself stand for the general tuning variable t that defines the region $\mathbf{Y}(a{=}act)$). With N different sensors there are, in fact, 2^N different actions, corresponding to which subset of the stations "choose to act." We suppose that all of the stations have identical impact on the cost function, so that the cost depends only on how many of the sensors act, and not on which ones they are. Specializing further to the case of two stations we see that the cost matrix will have three columns and two rows:

C(a,h)

Number acting	h=1	h=0
0	C(0,1)	C(0,0)
1	C(1,1)	C(1,0)
2	C(2,1)	C(2,0)

(101)

We may reasonably suppose that C(0,1) is the largest cost element, and C(0,0) is the lowest cost element. It is also clear that $C(j+1,0)>C(j,0)$. It is almost certainly the case that $C(0,1)>C(2,0)$. That is, the cost of the disease is greater than the cost of the cure. Any further assumptions are debatable, depending on the effectiveness of isolated action, the cost of resources consumed in responding, and so forth. For example, if a single response is totally effective then $C(1,1)=C(1,0)=$the cost of making one response, and $C(2,1)=C(2,0)>C(1,1)$ because further resources are consumed unnecessarily. When a single response is not certain to

be effective this inequality is likely to be reversed.

Let F_j stand for the false alarm rate corresponding to the tuning of the $j^{\underline{th}}$ sensor and D_j stand for the corresponding probability of detection. The overall expected cost of this tuning is then:

$$EC(F_1,F_2)=p_0^0\Big(C(0,0)\overline{F_1}\,\overline{F_2}+C(1,0)[F_1\overline{F_2}+\overline{F_1}F_2]+C(2,0)F_1F_2\Big)$$
$$+p_1^0\Big(C(0,1)\overline{D_1}\,\overline{D_2}+C(1,1)[D_1\overline{D_2}+\overline{D_1}D_2]+C(2,1)D_1D_2\Big) \tag{102}$$

The variables F_1 and F_2 are free to range independently over the unit square in F_1,F_2 space. Our formulation of this problem is quite different from the treatment in sections 1-8. We have not explicitly separated the problem into the determination of a doc and the selection of an optimal point. The expected cost function here directly involves both the operating characteristic and the priors and cost parameters. Thus there will be a separate problem to solve for each choice of the parameters. However, the problem of determining the two doc boundary functions represented by $D_1(F_1)$ and $D_2(F_2)$ can be solved once and for all. They can be incorporated as subroutines in an overall optimization program to find the minimum cost tuning. The Kuhn-Tucker conditions of this optimization problem are the coupled equations of [Tenney81].

An alternative way of thinking of this problem is to note that it is a specific restriction of the case of the unrestricted product of the sensors, with four possible actions. The restriction is that the trigger regions in the combined signal set must have a simple product form $\mathbf{Y}=\mathbf{Y}_1\times\mathbf{Y}_2$. (See also [Sadjadi].)

10. Summary and Conclusions.

We have seen that the concept of the doc — a convex set whose boundary, in the simple case of binary hypotheses is the familiar Receiver Operating Characteristic — provides a useful unifying foundation for the discussion of both discrete and continuous sensor systems in a common language. This is particularly important because continuous approximations to discrete systems are a convenient way of achieving mixed strategies (See Section 8) while discrete approximations to continuous situations represent the realities of signal binning. We have made every effort to develop the language and notation in a way that will survive transition to the case of more than two actions or more than two hypotheses. Although this complicates the discussion of some of the most familiar cases, we believe that it is worth the effort.

We have shown that the doc $\mathfrak{D}(S)$, and the set of its extreme points $\mathfrak{B}^{+}(S)$, represented by $D(F)$ provide all the information needed to solve any Bayesian problem, for either coordinated or team action. We have further introduced a powerful notation for the calculus of sensors, built on the two fundamental operations of full sensor product $S \otimes T$ and the M-fold restriction representing either messaging or action, $\mathfrak{R}(M)S$.

Using this machinery we have found a number of "negative results" contradicting certain plausible beliefs about basic properties of networks. Specifically, we have shown that:

(1) The fact that sensors are identical, and that their messages are combined in a symmetrical fashion at a fusion center does not imply that the best tuning for the individual sensors is symmetrical itself. An example is given in Section 5.1.

(2) When one sensor is definitely better than another (that is, its doc completely contains that of the poorer sensor) it is not necessarily the case that it is better to combine full information from the better sensor with a restricted message from the poorer. Nor is the reverse the case. An example is given in Section 5.3 in which the doc's for both possible architectures are calculated, and it is shown that neither contains the other.

(3) When signals from several sources are to be combined to determine an action there

may be discontinuities in the optimal tuning, corresponding to the fact that the tuning of the fusion center itself is a discrete selection from a set of several possible LOGICs.

(4) Even though the optimal tuning for each of the sensors in a network will be a deterministic tuning (corresponding to a point on the boundary of its own doc), it is not the case that deterministic tuning of the network as a whole is always optimal. We show, by example, that when there are resource limitations, as well as cost criteria, the best possible deterministic fusion system may still be suboptimal.

Our results confirm the view that the development of an optimal architecture based on distributed sensors is a difficult problem. We have shown, by example, how one may construct the boundary of the doc for a complex system and may, once an optimal tuning t^* for the whole network has been chosen, "climb back" up the structure to determine the optimal tunings of all of the constituent sensors.

There are two important directions for immediate exploration. One is to find the most efficient possible algorithms for implementing the sensor calculus. We have been using straightforward grid search to test various preconceptions about symmetry and dominance. Particularly because symmetrical solutions are not generally optimal, the calculation for large numbers of sensors may become prohibitive, unless better algorithms can be found.

The second important direction is to establish bounds on the magnitude of the suboptimality represented by our various examples. For example, if it could be shown that symmetric solutions are always within .5% of the optimal tunings in fusion, then it might, in many cases, be acceptable to use the suboptimal symmetrical tuning with substantial computational savings. Similarly, if it could be shown that one series arrangement is always superior to another to within a similar small difference, it might be acceptable to eliminate the opposite architecture from consideration.

REFERENCES AND LITERATURE CITED

Bar-Shalom, Y; Campo, L. The effect of common processor noise on the two-sensor fused-track convariance. IEEE Trans. Aerospace and Electronic Systems. vAES-22(6)p803:5 (Nov. 1986)

Birdsall, T.G. [19873]. The theory of signal Detectability. ROC curves and their character.Report 177. Cooley Electronics Lab. University of Michigan. 1973

Blackwell, D.,"On randomization in statistical games with k terminal actions," in Kuhn and Tucker, pp 183-188(1953).

Blankenbecler, R. and H. Partovi, Phys. Rev. Lett. 54, 373 (1985).

Blankenbecler, R. and P. B. Kantor, Distributed Detector Systems as a Problem in Optimization with Nonlinear Constraints, Submitted to IEEE Transactions on Aerospace and Electronic Systems (Appendix A of this report).

Bobrow, L. and M. Arbib, Discrete Mathematics: Applied Algebra for Computer and Information Science, W. B. Saunders Co., Philadelphia, Pa., 1974.

Burg, J. P.,"Maximum Entropy Spectral Analysis," Ph. D. Thesis in Geophysics, Stanford University (1975).

Chair, Z., Ph.D. dissertation in progress, Syracuse University, N.Y.

Chang, K-C.; Chong, C-Y; Bar-Shalom, Y. Joint Probabilistic Data Association in distributed Sensor Networks. IEEE Trans Automatic Control vAC-31(10)p889-897 (Oct 1986)

Cheeseman, P., Proc. of 85th Int. Joint Conf. on Art. Intell.,Karlsruhe, p. 198, Aug., 1983.

Chu, Peter L. A Nominaotr-Selector Two-Stage Signal Detection Scheme. Proceedings of the IEEE Transactions on Automatic Control. AC-32p233:239

Cline, T. B. and R. E. Larson,"Decision and Control in Large-Scale Systems Via Spatial Dynamic Programming," Tutorial: Distributed Control, R. E. Larson (Ed.), IEEE Computer Society, 1978.

Ekchian, L."Optimal Design of Distributed Detection Networks," MIT PHD. Thesis, 1982.

Ekchian, L. K., and Tenney, R. R. (1982). Detection networks. In Proceedings of the 21st IEEE Conference on Decision and Control, Orlando, Fla., 1982.

Ekchian L. K. and R.R. Tenney, "Recursive Solution of Distributed Detection/Communication Problems", Proc. of the 1983 American Control Conference, Vol. 3, San Francisco, CA, pp. 1338-1339.

Fejar, A. (1978). Combining techniques to improve security in automated entry control.

Gallager, R. G., Information Theory and Reliable Communication, Wiley and Sons, Inc., New York, 1968.

Gull, S. F. and G. J. Daniel, Nature 272, 686 (1978).

Ho, Y. C., and K. C. Chu,"Team Decision Theory and Information Structures in Optimal Control Problems - Part I,"IEEE Trans. on Auto. Control, Vol. AC-17, No. 1, February 1972, pp. 15-22.

Ho, Y. C., and Mitter, S.K. (1976). Directions in Large-Scale Systems. New York: Plenum Press, 1976.

Ho, Y. C., Kastner, M. P., and Wong, E. (1978). Teams, signaling, and information theory. IEEE Transaction on Automatic Control, AC-23 (1978), 305-311.

Ho, Y. C.,"Team Decision Theory and Information Structures", Proceedings of the IEEE, Vol. 68, No. 6, June 1980, pp. 664-654.

Ho, Y. C., and T. S. Chang,"Another Look at the Nonclassical Information Structure Problems," IEEE Trans. on Auto Control, June 1980, pp. 537-540.

Huang, K-Y; Fu, K-S. Decision Theoretic Approach for Classification of Ricker Wavelets and Detection of Seismic Anomalies. IEEE Transaction on Geosciences and Remote Sensing. GE-25 p..:..

Jaynes, E. T., Phys. Rev. 106,620 (1957), and Phys. Rev. 108,171 (1957).

Johnson, R. W. and J. E. Shore, IEEE trans, Acoust. Speech and Signal Processing ASSP31, 574 (1983).

Kantor, P. B., Information Technology 3, 88 (1984)."Maximum Entropy and the Optimal Design of Automated Information Retrieval Systems."

Kantor, P. B., and Lee, J. J."The Maximum Entropy Principle in Information Retrieval" Proceedings of the ACM-SIGIR Conference in Pisa, Italy September, 1986. Ed. F. Rubatti.

Kovattana, T. (1973). Theoretical analysis of intrusion alarm using two complementary sensors. Final Technical Report, Stanford Research Institute, Menlo Park, Calif., Feb. 1973.

Kuratran, B.,"Dynamic Two Person Two Objective Control Problems with Delayed Sharing Information Pattern," IEEE Trans on Auto Control, Vol. AC-22, August 1972, pp. 659-661.

Kushner, H. J., and Pacut, A. (1982). A simulation study of a decentralized detection problem. IEEE Transaction cn Automatic Control, 27,5 (Oct. 1982), 1116-1119.

Lauer, G. and N. R. Sandell, Jr.,"Distributed Detection of Known Signals in Correlated Noise," ALPHATECH Report (in preparation), Burlington, MA, March 1982.

Middleton, D. (1969). An Introduction to Statistical Communication Theory. New York: McGraw-Hill, 1969.

Ng, L.C.; Bar-Shalom, Y.Multi-sensor multi-targe delay time estimation. IEEE Trans Acoustics Speech and Signal Processing vASSP-34(4)p669:678 (Aug 1986)

Papavassilopoulos, G. P., and J. B. Cruz. (1979). Nonclassical control problems and Stackelberg games. IEEE Transactions on Automatic Control, AC-24 (1979), 155-166.

Reibman, AR; Nolte LW. Optimal Design and Performance of Distributed Sensors. IEEE Transactions on Aerospace and Electronics Systems AES-21 [1987] p 24:30.

Rietsch, E. J., Geophysics, 42, 489 (1977). Sandell, Jr., N. R. and M. Athans,"Solution of Some Nonclassical LOG Stochastic Decision Problems," IEEE Trans. on Auto. Control, AC-19, April 1974, pp. 108-116.

Sadjadi, F.A., "Hypothesis Testing in a Distributed Environment," IEEE Transactions on Aerospace and Electronic Systems 22, Vol 2, pp. 134-137 (1986).

Sandell, N. R., Varaiya, P., Athans, M., and Safanov, M. G. (1978). Survey of decentralized control methods for large-scale systems. IEEE Transactions on Automatic Control, AC-23 (1978). 108-128.

Sarma, VVS; Gopala, Rao KA. [1983] Decentralized Detection and Estimation in Distributed Sensor Systems. in Proceedings of the IEEE 1983 Conference on Cybernetics and Society v1p438:441

Smith, R., For a general set of papers covering many types of applications of this method of reasoning, see the proceedings of Laramie and Calgary meetings on Maximum Entropy (1982,83,84) Foundations of Physics, edited by R. Smith (Plenum, New York, 1985) (in press).

Srinivasan, R. Distributed radar detection theory. IEEE Proceedings V133 Pt F, No 1. Feb 1986 p55:60

Srinivasan, R; Sharma, P.; Malik, V. Distributed detection of Swerling targets. IEEE Proc. Vol 133 Pt F, No 7 (Dec 1986) p624:629

Stearns, S. D. (1982). Optimal detection using multiple sensors. In Proceedings of the 1982 Carnahan Conference on Crime Countermeasures (May 1982).

Stengel, D.N., Luenberger, D. G. Larson, R. E. and T.B. Cline,"A Descriptor Variable Approach to Modeling and Optimization of Large Scale Systems," Systems Control Inc., Palo Alto, CA., Feb. 1979.

Teneketzis, D. (1982). The decentralized quickest detection problem. In Prodeedings of the 21st IEEE Conference on Decision and Control, Orlando, Fla., 1982.

Teneketzis, D. (1982). The decentralized quickest detection problem. Presented at the 1982 IEEE International Large-Scale Systems Symposium, Virginia Beach, Va., Oct. 1982.

Tenney, R. R."Distributed Decision-Making Using a Distributed Model," LIDS-TH-938, MIT, PHD. Thesis, Sept. 1979.

Tenney, R. R. and Sandell, N. R.,"Detection With Distributed Sensors." IEEE Transactions on Aerospace and Electronic Systems 17, Vol. 4, pp. 501-509 (1981), and 17, Vol. 5, p. 736 (1981).

Tenney, R. R. and Sandell, N. R. Jr., IEEE-AES Vol 17 n4 (July 1981) p. 501-510.

Thompson, G. L.,"Signalling Strategies in n-Person Games," Contributions to the Theory of Games II, Princeton, pp. 267-276.

Van Trees, H. L. (1966). Detection, Estimation and Modulation Theory. New York: Wiley, 1968.

Van Trees, H. L. (1969). Detection, Estimation and Modulation Theory, vol. I. New York: Wiley, 1969.

Varaiya, P. and J. Walrand,"On Delayed Sharing Patterns," IEEE Trans. on Auto. Control, Vol.23, June 1978, pp. 388-394.Witsenhausen, H. S.,"A Counterexample in Stochastic Optimum Control," SIAM, Vol. 6, No. 1, 1968, pp. 138-147.

Witsenhausen, H. S.,"A Standard Form for Sequential Stochastic control," Math. Sys. Th., Vol. 7, 1973.

SLAC – PUB – 4510
December 1987
(N)

DISTRIBUTED DETECTOR SYSTEMS AS A PROBLEM IN OPTIMIZATION WITH NONLINEAR CONSTRAINTS

R. BLANKENBECLER*
Stanford Linear Accelerator Center,
Stanford University, Stanford, California 94305
and
PAUL B. KANTOR†
Tantalus, Inc., 3257 Ormond Road, Cleveland, Ohio 44118

Summary

In this paper, optimal control theory is applied to the design of decentralized sensor systems. Lagrange inequality multipliers are used to determine the optimal design parameters. Several models of possible response functions are fully discussed as examples of our technique.

1. Introduction and Definition of the Problem

There are many situations in science and engineering in which information is gathered from a variety of sensors and must be abstracted or summarized for future processing in order to comply with communication, storage, or processing constraints. The simplest example is the case in which a binary decision must be made based upon information sources that are constrained to transmit a binary signal. Examples include data from devices monitoring the performance of a power network, data from an array of elementary particle detectors, the coordination of radar or infrared signals, and so on.

In general, the communication restrictions may be lifted with some increase in cost; thus the examples under discussion represent a special case. As we shall see, even this simple case (two alternative states, two possible actions, two-fold signals, and two detectors) presents challenging problems of analysis. Discussions have been given by Srinivasan for more than two detectors[1] and with applications to a specific choice of the detector characteristics.[2]

Discussion of a case with distributed action is given by Tenney and Sandell.[3] A discussion for specific (series) topologies is given by Ekchian and Tenneyl.[4] Related problems have been discussed by Chair and Varshney,[5] by Reibman and Nolte,[6] and by Sadjadi.[7]

Quite generally, the performance of an entire network is summarized by four probabilities $p_r(y, H)$, of which only two are independent. (Here, $H = H_0, H_1$ represents two hypothesis about the world and $y = y_0, y_1$ represents two possible actions or determinations. This notation will be made more precise shortly.)

Several problems may be formulated, including

(i) min $p_r(y_1, H_0)$ subject to $p_r(y_0, H_1) \le p_m^0$.

(ii) min $p_r(y_0, H_1)$ subject to $p_r(y_1, H_0) \le p_f^0$.

(iii) min $Ap_r(y_1, H_0) + Bp_r(y_0, H_1)$.

The first and second problems correspond to setting acceptable error rates; the third arises when there is a tradeoff between the two types of error. The coefficients A, B may be positive or negative.

The physical characteristics of an individual detector constrain the achievable values of $p_r(y_1, H_0)$ and $p_r(y_0, H_1)$. The design of a network is then a selection from among a discrete set of topologies, with each topology tuned to give its best possible performance. The tuning is a constrained optimization, with the constraints determined by the achievable values of $p_r(y_1, H_0)$ and $p_r(y_0, H_1)$.

*Work supported by the Department of Energy, contract DE-AC03-76SF00515.
†Supported in part by the Office of Naval Research, contract N00014-87-C-0695.

Our work has many points of contact with previous work. We utilize a Lagrangian formulation to deal with the optimization problem involving equality and inequality constraints. Three problems are presented in detail, involving the cases of exponential response functions, special sums of exponentials, and block functions. We trace the behavior of the system tuning and the optimal cost as a function of the detector discrimination.

This paper is the first of a series whose goal is to clarify the relations between topics in distributed detection, optimal control, and experimental design, thereby leading to a more intuitive or "physical" understanding of the problems of distributed detection and sensing.

1.1 General Introduction

There are two possible states (of the world) H_0 and H_1. The prior probabilities of these two states are p_0 and p_1, where

$$\begin{aligned} p_0 &= p_{rior}(H_0) \\ p_1 &= p_{rior}(H_1) . \end{aligned} \tag{1}$$

There are two possible courses of action ("measures") denoted by m_0 and m_1.

The assumed *cost* function is $C(m, H)$, where

$$\begin{aligned} C(m_0, H_0) &= u_0 \qquad & C(m_0, H_1) &= u_1 + w_{01} \\ C(m_1, H_0) &= u_0 + w_{10} \qquad & C(m_1, H_1) &= u_1 . \end{aligned} \tag{2}$$

The expectation value of the cost function is to be minimized over the various design parameters, those in the response functions and those in the probability functions. As will become clear later in our discussion, the separate cost parameters u_0 and u_1 do not matter when the expected cost is minimized; the minimum depends only on a ratio involving the differences in the cost for a given H_j, namely w_{10} and w_{01}.

The essential point is that for the case of only two possible states of the world, the preferred action is determined by a single real number, determined by the posterior odds for the H_j. This is true because, using linear cost theory, the information in Eq. (2) is summarized by the intersection point of two straight lines; one describes the cost of action m_0 as a function of p_0 while the other describes the cost of action m_1.

1.2 Properties of the Integrator

For our model we choose a fusion structure in which signals are processed locally at each detector, with messages fed to a single integrator

$$A \longrightarrow C \longleftarrow B .$$

The problem is to design an integrator C and tune the sensors (A, B). Each of the two sensors detects some signal (y) and sends the central integrator a signal u_i. In general, these signals need not be binary. The integrator then chooses action m_0 or m_1, and this choice is determined by the fusion rules. The rules for both the sensors and the integrator are to be chosen so that the expected cost is minimized.

The integrator's actions are completely described by a matrix (with two adjustable parameters) that describes the probability of choosing measure m_0, given the signals u_i from detector $i = a, b$. This matrix will be denoted by $p(m_i|u_a, u_b)$, where

Submitted to Transactions on Aerospace and Electronic Systems

$$p(m_0|0,0) = 1 \qquad p(m_0|0,1) = g$$
$$p(m_0|1,0) = d \qquad p(m_0|1,1) = 0 \tag{3}$$

or, in an alternative matrix notation,

$$p(m_0|\vec{u}) = \begin{array}{c|cc} & u_b = 0 & u_b = 1 \\ u_a = 0 & 1 & g \\ u_a = 1 & d & 0 \end{array} . \tag{4}$$

The probability of choosing m_1 must be the complement, element by element,

$$p(m_1|\vec{u}) = 1 - p(m_0|\vec{u}) . \tag{5}$$

We exclude the possibility of a third course of action. The design parameters g and d are to be fixed by the optimization; they define the rule to follow when the two detectors disagree. If the two detectors are identical, then we expect that $d = g$ and that they will be 0 or 1 depending on the costs and the details of the sensitivities of the detectors.

1.3 Definition of the Detectors

Now consider the detectors in more detail. Each detector, labeled a or b, produces a single "meter reading" y_i, $(i = a, b)$, in response to the state of nature. The probabilities, $p_d(y, H)$, that the value of the reading is y for the state of the environment H for each detector is

Detector→	a	b
$p_d(y; H_0)$	$f_0^a(y)$	$f_0^b(y)$
$p_d(y; H_1)$	$f_1^a(y)$	$f_1^b(y)$.

The quantity y must now be converted to a yes or no signal ($u = 0$ or $u = 1$).

The effect of the decision process at each detector may be summarized *completely* by a table giving the decision strategy or probability of response $p_i(u; H)$ for each of the detectors. For detector a:

$$p_a(u_a; H) = \begin{array}{c|cc} & H_0 & H_1 \\ u_a = 0 & a & 1 - A \\ u_a = 1 & 1 - a & A \end{array} , \tag{6}$$

while for detector b:

$$p_b(u_b; H) = \begin{array}{c|cc} & H_0 & H_1 \\ u_b = 0 & b & 1 - B \\ u_b = 1 & 1 - b & B \end{array} . \tag{7}$$

1.4 Design Parameters

Therefore, the full set of parameters to be determined by optimization is

$$g, d \qquad a, A, b, B .$$

The first two describe the operation of the "integrator" that processes the two signals of the sensor stations to form the output decision m. The last four describe the operation of the "sensors" — they take the detected signals, apply their respective detection criteria, and form their individual output signals $\vec{u}$.

1.5 Properties of the Detectors

A generalized detector uses the rule: if the signal y is in the region R, then the signal u_0 is sent to the integrator. Similarly, if the signal is in the complement of R, i.e., if $y \in \bar{R}$, then u_1 is sent.

If the external state is indeed H_0, then the response function of the detector is $f_0(y)$, but if it is H_1, the detector responds with $f_1(y)$ [see the table below Eq. (5)].

For any choice of R, the detection probability [see Eq. (6)] is

$$a = \int_R dy\, f_0^a(y) , \tag{8}$$

and this then implies for A,

$$A = \int_{\bar{R}} dy\, f_1^a(y) . \tag{9}$$

Similar relations hold for b and B. If the response functions have interlaced maxima, then the region R (and $\bar{R}$) may be disconnected.

As R expands, clearly $\bar{R}$ contracts. For any fixed value of a there is a maximum and a minimum possible value for A. If the response functions $f_0(y)$ and $f_1(y)$ overlap, which is the general and expected case, then these limits on the value of A have important consequences.

The possible values A for a fixed value of a, are traversed as the region R is varied. It is clear that to make A as large as possible for a given value of a, R should contain those points whose contribution to A would be as small as possible (i.e., the ratio f_1/f_0 small) while the complement contains those points with large values of this ratio. This is the familiar likelihood ratio threshold rule.

If a goes to 1, then A goes to zero. Also, if A is 1, then a must vanish. This follows trivially from the unit normalization of the response functions.

Finally, note that an ideal detector with perfect discrimination has response functions that satisfy $f_0(y) \times f_1(y) = 0$ for *all* y. In this case, the values of a and A are independent. We will return to this limiting case shortly.

2. The Cost Function

The expected value of the cost function is

$$\langle C \rangle = \sum_{i,j} C(m_i, H_j) p_{prior}(H_j) p(m_i|H_j) , \tag{10}$$

where $p(m|H)$ is directly expressed in terms of the detector properties, and we assume that the signals received by the detectors are stochastically independent:

$$p(m_i|H_j) = \sum_{u_a, u_b} p(m_i|u_a, u_b) p(u_a|H_j)\, p(u_b|H_j) . \tag{11}$$

Using the explicit form of the cost matrix, Eq. (2), (10) can be expressed as

$$\langle C \rangle = w_{01}\; p(m_0|H_1)\, p_1 + w_{10}\; p(m_1|H_0)\, p_0 + u_0\, p_0 + u_1\, p_1 . \tag{12}$$

Additive constants do not matter in the minimization; the last two terms are fixed, and are the cost for an ideal system. For such a system with perfect discrimination, the off-diagonal probabilities $p(m_0|H_1)$ and $p(m_1|H_0)$ both vanish since $A = 1 - a$. The cost must be a minimum:

$$\langle C \rangle_{min} = u_0\, p_0 + u_1\, p_1 . \tag{13}$$

The quantity that we want to minimize is the additional cost due to imperfections in the system; this has the form

$$\begin{aligned} \langle \delta C \rangle &= \langle C \rangle - \langle C \rangle_{min} \\ &= w_{01}\; p(m_0|H_1)\, p_1 + w_{10}\; p(m_1|H_0)\, p_0 . \end{aligned} \tag{14}$$

Note that the position of the minimum will depend on the ratio

$$W = \frac{w_{10}\, p_0}{w_{01}\, p_1} , \tag{15}$$

which is the relative expected cost of being wrong if the state of the environment is H_0 (and responding with m_1) compared to the cost of being wrong if it is H_1 (and responding with m_0). The magnitude of the minimum cost will depend multiplicatively on the factor $w_{01}\, p_1$.

It is convenient to rewrite the cost function as

$$J \equiv \langle \delta C \rangle / (w_{01}\, p_1) , \tag{16}$$

or

$$J = p(m_0|H_1) + W[1 - p(m_0|H_0)] . \tag{17}$$

The minimization of the expected value of the cost function is equivalent to minimizing J.

Some interesting limits on J can now be determined. The perfect detector has $J = 0$. *It is amusing to note that a detector* that is *always* wrong has $J = 1 + W$. (One would then use such a detector "backward.") A more interesting case follows from noting that if W is sufficiently small, i.e., the cost w_{10} (of erroneously choosing m_1) is small, then a good strategy is to always choose m_1. This implies that $p(m_0|H_1) = p(m_0|H_0) = 0$, and $J = W$ (and $g = d = 0$). If, on the other hand, W is larger than 1, then one wants to always choose m_0; in this limit, $J = 1$ (and $g = d = 1$). The final cost for this limiting case may be expressed in terms of the step function $\theta(x)$ ($\theta(x) = 1, x > 0, \theta(x) = 0, x < 0$)

$$\begin{aligned} J_{max} &= W\theta(1-W) + \theta(W-1) \\ g &= d = \theta(W-1) . \end{aligned} \tag{18}$$

This result arises in another way. If the response functions are the same, $f_0(y) = f_1(y)$, then no discrimination is possible, and we find $A = 1 - a$. Using this relation in the probabilities, we find the above result by choosing the obvious optimum.

The general optimization problem consists of choosing the design parameters so the expected cost lies as far below J_{max} as possible and as close to the ideal case, $J = 0$, as possible. We now turn to a general discussion of the problem of finding extrema when the constraints define a connected subset of the real line for each variable.

3. General Minimization with Inequalities

Using the form of the probabilities defined in Eqs. (6) and (7), one finds the explicit expressions

$$\begin{aligned} p(m_0|H_1) &= 1 - p(m_1|H_1) , \\ &= (1-A)(1-B) + g(1-A)B + dA(1-B) , \end{aligned} \tag{19}$$

and

$$\begin{aligned} p(m_0|H_0) &= 1 - p(m_1|H_0) , \\ &= ab + g\,a(1-b) + d(1-a)b . \end{aligned} \tag{20}$$

Using Eqs. (19) and (20), the minimization problem can be re-cast explicitly as

$$J = W + [S + g\,T + d\,U] , \tag{21}$$

where

$$\begin{aligned} S &= (1-A)(1-B) - Wab \\ T &= (1-A)B - Wa(1-b) \\ U &= A(1-B) - W(1-a)b , \end{aligned} \tag{22}$$

and all the variables must satisfy inequality constraints. A complete mathematical treatment for problems of this type can be found in the excellent book by Hestenes.[8] A reference that discusses such variational problems in a language perhaps more familiar to physicists and engineers is available.[9]

To *minimize* J, in the case that the variables g and d occur linearly in J, but have a restricted range from zero to one, it is convenient to form the variational functional J_{var}, where

$$J_{var} = J - \gamma g(1-g) - \delta d(1-d) . \tag{23}$$

The optimum will be a saddle point in (γ, δ) versus (g, d). In this case J is a linear function of g and d, hence the extrema will occur at the endpoints. The Lagrange inequality multipliers γ and δ must be zero if their associated variable g or d is inside the allowed range, and non-negative if they are on the boundary.[9] As usual, the derivative with respect to g must vanish at the minimum and this yields the condition

$$0 = T - \gamma(1-2g) . \tag{24}$$

This takes the place of paired Kuhn-Tucker conditions for $g \geq 0$ and $g \leq 1$. If T is nonzero, which is the typical case, then the minimum must be on the boundary (γ cannot be zero). If T is positive, then g vanishes; if negative, then g is unity. A similar argument holds for d and U. The result can be expressed as

$$\begin{aligned} g &= \theta(-T) \\ d &= \theta(-U) , \end{aligned} \tag{25}$$

and the minimum of J_{var} becomes

$$J_m = W + [S + T\theta(-T) + U\theta(-U)] . \tag{26}$$

Note that if T or U vanish, there is no uncertainty in the minimum of J, even though g and d are not determined.

The variables left to consider are a, A, and b, B. Each of these variables has a restricted range, so inequality multipliers will again be used. As was noted before, the possible values of A are limited by the form of the response function and the value of a. This can be expressed as the statement that for any choice of the region R, with a given by (8), one must have

$$A_{min}(a) \leq A(R) \leq A_{max}(a) . \tag{27}$$

Of course, similar restrictions apply to B.

These inequalities can be treated as above. Write the variational functional in the form

$$J_{var} = J_m - F - f , \tag{28}$$

where

$$F \equiv \alpha_A (A - A_{min})(A_{max} - A) + \beta_B (B - B_{min})(B_{max} - B) , \tag{29}$$

and

$$f \equiv \alpha a(1-a) + \beta b(1-b) . \tag{30}$$

Again, the Lagrange inequality multipliers $\alpha_A, \beta_B, \alpha$ and β must be zero if their associated variable is inside the allowed range and non-negative if they are on the boundary.

Now the variation with respect to A yields

$$2\alpha_A (A - A_{bar}) = -\frac{\partial J_m}{\partial A} , \tag{31}$$

where $A_{bar} \equiv (A_{min} + A_{max})/2$.

It is a straightforward task, though somewhat tedious, to discuss the general case. First note that the above equation becomes

$$2\alpha_A (A - A_{bar}) = +(1-B) + B\theta(-T) - (1-B)\theta(-U) . \tag{32}$$

Since the right-hand side is never negative, α_A cannot vanish, and hence A must be at its boundary. Since α_A must also be non-negative, it follows that A must be above A_{bar}. Repeating the same argument for B we find that

$$\begin{aligned} A &= A_{max}(a) , \\ B &= B_{max}(b) . \end{aligned} \tag{33}$$

These are computable functions of a and b given the response functions of the detectors. They correspond to the so-called Receiver Operating Characteristic used in several of the papers cited above. We shall term these functions the DOC, or Detector Operating Characteristic, and they will play a fundamental role in our analysis.

The next stage is to vary a and b within their allowed range to achieve the overall minimum. One can anticipate that there may be symmetric ($a = b$) and nonsymmetric minima; which particular one is the global minima must be determined from a more detailed examination using the explicit forms for the response functions. This will be carried out in the explicit examples discussed in the next section. First let us discuss the boundary behavior in a and b.

<u>*Double boundary:*</u> The boundary region in which both variables are at their limits consists of four terms. They will be denoted by *L(a,b)*, where a and b can take on the values zero or one.

<u>*L(0,0):*</u> For this case, $A = B = 1$, and $J_m = W$ for all W.

$\underline{L(1,0) \text{ and } L(0,1)}$: For these cases, $A = 0\, B = 1$, or the reverse, and $[S = 0 = U, T = 1 - W$ and $g = \theta(W-1)]$

$$J_m = W + (1-W)\,\theta(W-1)\,, \tag{34}$$

the J_{max} discussed earlier.

$\underline{L(1,1)}$: For this limit, $A = B = 0$,

$$J_m = 1\,. \tag{35}$$

Therefore, the minimum of J_m on this double boundary is *always* given by Eq. (34) which amounts to setting the detectors to always signal oppositely. Now let us turn to the single variable boundaries.

Single boundary: This boundary region is symmetric in both variables and hence we need only treat the case in which b is at its limits while a is in the interior. The reversed situation will yield the same minima. These will be denoted as:

$\underline{L(a,0)}$: For this case, $B = 1$, and $S = U = 0$. The quantity T is not zero, with

$$\begin{aligned} J_m &= W + T\,\theta(-T) \\ T &= 1 - A - Wa\,. \end{aligned} \tag{36}$$

As noted, A should be equal to its maximum value for a fixed value of a in order to achieve the minimum value of T, as was shown earlier. The limit cases of $a = 0$ and $a = 1$ are on the double boundary. Any minimum for a in the interior must satisfy

$$\begin{aligned} \frac{\partial T(a)}{\partial a} &= \frac{\partial A_{max}(a)}{\partial a} - W \\ &= 0\,. \end{aligned} \tag{37}$$

Since $A_{max}(a)$ is a decreasing function of a, there will in general be a solution in this region if W is in an appropriate range. This could yield a smaller minimum than that given by the double boundary result, Eq. (34); however, for this case, we have $g = 1$ and d is not determined, but its value does not matter since $U = 0$ and one can arbitrarily choose $d = 1$ also.

$\underline{L(a,1)}$: For this situation, $B = 0 = T$ and $S = 1 - A - Wa$, with $U = 1 - W - S$. If U is positive, then $J_m = 1$, while if it is negative, then $J_m = W + S$. Both these cases have arisen before, and there are no new minima of J_m.

Interior: In the interior region, the inequality multipliers must vanish and the standard variational equations become symmetric in form. One can safely assume that there will be minima in this region, but whether any is the global minimum requires detailed study. Note that generally there will be (local) minima with T(and/or U) both positive and negative with the corresponding limiting values of g (and d). Since this is a standard well-discussed variational problem, further general treatment here is not necessary. Let us now turn to a exhaustive discussion of some explicit examples.

4. Exponential Response Functions

Consider a detector with response functions given by

$$\begin{aligned} f_0 &= n\lambda \exp\{-n\lambda y\} \\ f_1 &= \lambda \exp\{-\lambda y\}\,. \end{aligned} \tag{38}$$

We will assume that n is greater than one without any loss of generality, so that the likelihood ratio (f_1/f_0) is less than one for $y \le x$, where $\lambda x = (\ell n\, n)/(n-1)$. Using the above argument, to achieve the extrema of A for these monotonic response functions, the region R must be either the range below or the range above some point x whose value will be determined by the optimization process.[10]

Therefore, it is easy to see that there are two cases to discuss:

	Case I	*Case II*
R	$0 \le y \le x$	$x \le y \le \infty$
$\bar{R}$	$x \le y \le \infty$	$0 \le y \le x$
a	$1 - \exp\{-n\lambda x\}$	$\exp\{-n\lambda x\}$
A	$\exp\{-\lambda x\}$	$1 - \exp\{-\lambda x\}$
A	$(1-a)^{1/n}$	$1 - a^{1/n}$.

Thus, A must lie in the region

$$1 - a^{1/n} \le A \le (1-a)^{1/n}\,, \tag{39}$$

and its position in this interval is determined by the particular choice of the region R. Similar relations hold for b and B.

As an example, consider the case $n = 2$, and then the feasible region for A as a function of a is labeled F in the graph shown in Fig. 1.

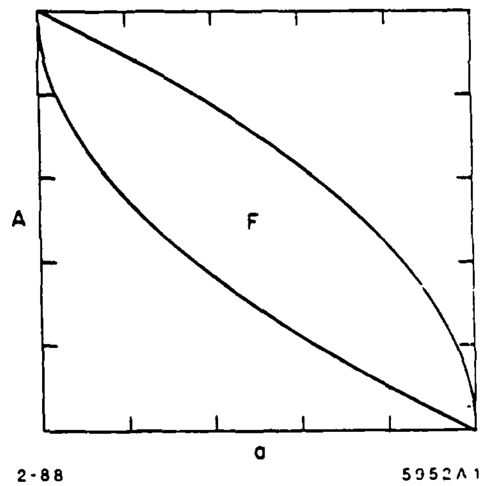


Fig. 1. *The allowed region of* A *is plotted for* $n = 2$ *as a function of* a *and labeled* F.

Figure 2 shows the graph for the value $n = 4$.

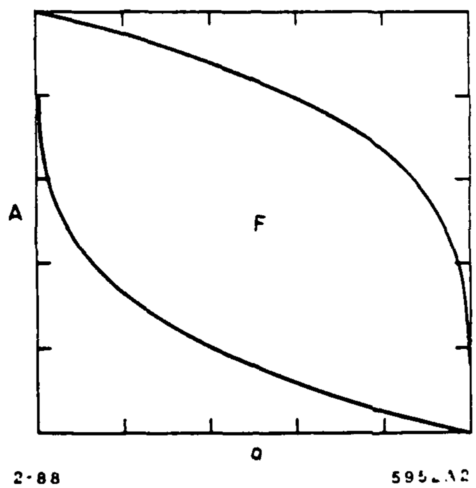


Fig. 2. Same as Fig. 1 but with $n = 4$.

We see that as n increases, the allowed region increases to eventually include all values of A between zero and one.

$\underline{4.1 \ \textit{Explicit Minimization}}$

Using the general results derived in the previous chapter, we have A and B at their maximum allowed values:

$$\begin{aligned} A &= (1-a)^{1/n} \\ B &= (1-b)^{1/n}\,. \end{aligned} \tag{40}$$

find no point where J was below the value at the symmetric minimum given above.

Note that as a function of $W \equiv w_{10}\,p_0/w_{01}\,p_1$, the largest fractional improvement in cost is achievable when $W = 1$. This is precisely the case in which the *prior* choice of action is a matter of indifference, that is,

$$u_0\,p_0 + u_1\,p_1 + w_{01}\,p_1 \;=\; u_0\,p_0 + w_{10}\,p_0 + u_1\,p_1 \;. \qquad (54)$$

This is intuitively reasonable, as one expects the information from the sensor to be the most valuable in this case.

5. Invariant Imbedding

We now consider a detector whose response functions allow superior discrimination between the two possible states of the environment and contains the previous example as a special case. The general form that allows a smooth limit back to the previous model is

$$\begin{aligned} f_0 &= n\lambda\,\exp\{-n\lambda y\} \\ f_1 &= \frac{n+1}{n+1-z}\lambda\,\exp\{-\lambda y\}\,\left[1 - z\,\exp\{-n\lambda y\}\right] \,. \end{aligned} \qquad (55)$$

For values of z near 1 this allows the improved separation between f_0 and f_1 since the former is large at $y = 0$ while the latter is small there. On the other hand, for z equal to zero, this is the model of the previous section.

It will again be assumed that n is greater than 1, and proceeding as before we find:

	Case I	*Case II*
R	$0 \le y \le z$	$z \le y \le \infty$
$\bar{R}$	$z \le y \le \infty$	$0 \le y \le z$
a	$1 - \exp\{-n\lambda z\}$	$\exp\{-n\lambda z\}$
A	$\frac{1}{n+1-z}\exp\{-\lambda z\} \times [n+1-z\exp\{-n\lambda z\}]$	$1 - \frac{1}{n+1-z}\exp\{-\lambda z\} \times [n+1-z\exp\{-n\lambda z\}]$
A	$(1-a)^{1/n}\left[1 + \frac{za}{n+1-z}\right]$	$1 - a^{1/n}\left[1 + \frac{z(1-a)}{n+1-z}\right]$.

Thus, A must lie in the region

$$1 - a^{1/n}\left[1 + \frac{z(1-a)}{n+1-z}\right] \le A \le (1-a)^{1/n}\left[1 + \frac{za}{n+1-z}\right] \,. \qquad (56)$$

Its position in this interval is determined by the particular choice of the region R. Similar relations hold for b and B. Note that the allowed region of A increases as z increases from zero to one.

To provide maximum contrast with the previous model we will present data for the value $z = 1$. For this case, the interior symmetric minimum exists for all W values. An interesting new behavior is found in this model for small enough W and n; the minimum cost occurs for $g = d = 0$, as before, but as W increases, these design parameters flip to $g = d = 1$. The value of a at the minimum jumps *discontinuously*.

	$n = 2$			$n = 4$		
W	a	J	J/J_{max}	a	J	J/J_{max}
0.1	0.46	0.088	.88	0.70	0.068	.68
0.2	0.58	0.160	.80	0.80	0.111	.56
0.25	0.63	0.191	.764	0.82	0.128	.51
0.26	0.64	0.197	.758	0.828	0.131	.504
0.27	0.24	0.203	.752	0.832	0.134	.496
0.3	0.26	0.22	.733	0.84	0.143	.48
0.5	0.35	0.31	.62	0.89	0.191	.38
1.0	0.50	0.47	.47	0.94	0.268	.27
1.5	0.60	0.56	.56	0.96	0.316	.32
2.0	0.68	0.63	.63	0.97	0.35	.35
4.0	0.79	0.77	.77	0.985	0.44	.44
6.0	0.85	0.83	.83	0.990	0.49	.49
8.0	0.88	0.87	.87	0.993	0.52	.52
10.0	0.90	0.89	.89	0.994	0.55	.55

The columns are the same as in the previous table. Note that the parameters for the minimum (for $n = 2$) shows a definite jump as W passes through the value ≈ 0.265. At this point, the optimum values of g and d change from zero to one; in fact, we find that $g = d = \theta(W - W_0)$, where $W_0 \approx 0.265$.

On the other hand, for $n = 4$, the quantities T and U are always positive, so that $g = d = 0$. There does not appear to be a discontinuity in a.

At the discontinuity, the *cost* varies smoothly. This is intuitively reasonable, since cost is, ultimately, determined by the position of some tangent hyperplane, along a normal to the feasible region, which is connected. However, the jump in design parameters could have serious consequences because a small variation in the (frequently subjective) data summarized by the parameter W could require a complete change of the system parameters g and d. This phenomena has important implications for the design of constant false alarm rate systems, which will be discussed elsewhere.[11]

6. A Step Function Example

We now consider a detector whose response functions, in a certain limit, allow a clean discrimination between the two possible states of the environment. In that limit, $A \to 1$ does not force a to zero. We will assume simple "square" response functions for ease of presentation. The response functions are chosen to be zero for $y > 3$ and, of course, normalized.

Proceeding as before we find for Case I, $0 < y < x$:

For the range	$0 \le x \le 1$	$1 \le x \le 2$	$2 \le x \le 3$
$f_0(x)$	$(1-\lambda_0)$	λ_0	0
$f_1(x)$	0	λ_1	$(1-\lambda_1)$
a	$(1-\lambda_0)x$	$1+\lambda_0(x-2)$	1
A	1	$1-\lambda_1(x-1)$	$(1-\lambda_1)(3-x)$.

A similar table can be evaluated for Case II, $x < y < 3$. If either λ_0 or λ_1 vanish, then this describes an ideal detector system.

We need the value of A_{max} for a fixed value of a which is

$$A_{max} = 1 - \frac{\lambda_1}{\lambda_0}(a - 1 + \lambda_0)\theta(a - 1 + \lambda_0) \,, \qquad (57)$$

while the minimum value is

$$A_{min} = \frac{\lambda_1}{\lambda_0}(\lambda_0 - a)\theta(\lambda_0 - a) \,. \qquad (58)$$

The feasible region for A as a function of a , labeled F in the graph (Fig. 3), is bounded by straight lines:

In the limit that either λ_0 or λ_1 vanishes, the allowed region for A covers the unit square.

It is a simple matter to analyze this problem for the minimum J corresponding to the maximum allowed A and B values as given above. Consider the cases:

1. $A = B = 1$, and $a = b = (1 - \lambda_0)$. For these values, T and U are negative and $J = W\lambda_0^2$.
2. $A = B = (1 - \lambda_1)$ and $a = b = 1$. For this case T and U are now positive and $J = \lambda_1^2$.

Thus the final result can be expressed as

$$\begin{aligned} J &= \min[W\lambda_0^2 \,,\ \lambda_1^2] \\ g &= d = \theta(\lambda_1^2 - \lambda_0^2 W) \\ a &= b = 1 - \lambda_0\,\theta(\lambda_1^2 - \lambda_0^2 W) \\ A &= B = 1 - \lambda_1\,\theta(\lambda_0^2 W - \lambda_1^2) \,. \end{aligned} \qquad (59)$$

The limit of perfect discrimination, λ_0 and/or λ_1 going to zero, can be easily discussed from the above results.

Varying J_m with respect to a and b and introducing inequality multipliers α and β to keep these variables between zero and one, we find the conditions

$$\begin{aligned} n\alpha(1-2a) &= A^{1-n}(1-B) - nWb \\ n\beta(1-2b) &= B^{1-n}(1-A) - nWa \end{aligned} \tag{41}$$

whose solution should contain all relevant minima. Let us examine the boundary and interior minima in that order. Recall that $n \geq 1$ in the following discussion, and we have assumed for the moment that T and U are positive. This will be proven shortly for our solutions.

<u>Boundary:</u> The double boundary region has been discussed in general and the result is a minimum of the form ($a = 0, b = 1$ or $a = 1, b = 0$)

$$J_m = W + (1-W)\,\theta(1-W) \equiv J_{max}\,. \tag{42}$$

<u>L(a,0):</u> For this single boundary problem, the task is to find the minimum of T, where $T = 1 - A_{max} - Wa$. The result is with $p = 1/(n-1)$

$$\begin{aligned} A_0 &= (\frac{1}{nW})^p \\ a_0 &= 1 - \frac{A_0}{nW}\,. \end{aligned} \tag{43}$$

For A_0 to be less than one, $nW \geq 1$. The value of T at this minimum is negative, and

$$J_m(Bnd) = 1 - \frac{n-1}{n}\,A_0\,. \tag{44}$$

If nW is slightly larger than one, $nW = 1 + \epsilon$, then it is easy to see that to lowest order

$$J_m(Bnd) \sim J_{max} - \frac{\epsilon^2}{2(n-1)}\,. \tag{45}$$

<u>L(a,1):</u> For this case, $B = 0 = T$ and $U = 1 - W - S$. If U is negative, then the minimum of J_m is 1. If it is positive, then S must be minimized, and this is just the problem discussed above.

In summary, J_m has a minimum on the boundary given by Eq. (42) or Eq. (44), depending on the value of nW.

<u>Interior:</u> In the interior region, the inequality multipliers α and β must vanish and Eqs. (41) become symmetric in form. Thus there is a symmetric solution with $a = b$ and $A = B$. Unsymmetric solutions will be searched for later. In the symmetric case, the equation for the optimal probability A is

$$nWA^{n-1}(1-A^n) = (1-A)\,, \tag{46}$$

which does not have an analytic solution for general n. The limiting behavior of the solution is easily extracted. For large W, A approaches zero, and a approaches one with the behavior [recall that $p = 1/(n-1)$]

$$\begin{aligned} A &\simeq \left(\frac{1}{nW}\right)^p + \ldots\,, \\ a &\simeq 1 - \left(\frac{1}{nW}\right)^{1+p} + \ldots\,. \end{aligned} \tag{47}$$

This is similar to one of the boundary solutions. The minimum of J in this limit has the form

$$J \simeq 1 - 2\frac{n}{n-1}A + \ldots\,. \tag{48}$$

Let us now discuss small values of W. Note that as W decreases, a decreases. The value of W where a vanishes is

$$W(a=0) = 1/n^2\,. \tag{49}$$

For values of W smaller than this value, there is no interior symmetric solution. At this critical value, $S = 0$. Finally, note that for this symmetric solution, $T = U$, and using the above equations,

$$T = (n-1)WaA^n\,, \tag{50}$$

which is positive definite. Therefore, the T and U terms do not contribute to this minimum because $g = d = 0$.

Using the equation for A, we find at the minimum

$$J_m(Int) = W + (1-A)^2 - W(1-A^n)^2\,. \tag{51}$$

This is smaller than the minimum arising from the boundary.

To see this, study the difference of Eq. (44) and Eq. (51) for sufficiently large W (so that the former exists). If W is eliminated between Eqs. (46) and (43), the result is

$$A_0 = A\,(\frac{1-A^n}{1-A})^p\,, \tag{52}$$

which shows that $A_0 = A_0(A) \geq A$. The difference becomes

$$\begin{aligned} J_m(Bnd) - J_m(Int) &= [1-(1-A)^2] - (\frac{1}{n})A_0^{1-n} \\ &\quad [1-(1-A^n)^2] - \frac{n-1}{n}A_0\,. \end{aligned} \tag{53}$$

For large W this difference approaches zero as $(n-1)/(n^2W)$. For all values of nW larger than one it is a simple matter to show that it is positive (a numerical proof is easiest).

Some sample numerical results are:

	$n=2$		$n=4$	
n^2W	a	J-W	a	J-W
0 – 1	0.00	-0.000	0.00	-0.00
1.0	0.00	-0.000	0.00	-0.00
1.1	0.120	-0.0001	0.082	-0.0000
1.2	0.218	-0.0009	0.150	-0.0001
1.3	0.299	-0.0026	0.210	-0.0003
1.4	0.367	-0.0054	0.261	-0.0007
1.6	0.475	-0.0144	0.346	-0.0018
1.8	0.556	-0.0278	0.413	-0.0036
2.0	0.618	-0.0451	0.467	-0.0061
3.0	0.791	-0.175	0.636	-0.0260
4.0	0.866	-0.348	0.725	-0.091
6.0	0.930	-0.757	0.815	-0.131
8.0	0.957	-1.203	0.862	-0.219
10.0	0.971	-1.669	0.890	-0.315
12.0	0.979	-2.144	0.909	-0.417
16.0	0.987	-3.117	0.933	-0.629

Recall that $g = d = 0$ for this global minimum. Therefore, if either detector signals 1, one should make the choice m_1 for any value of W.

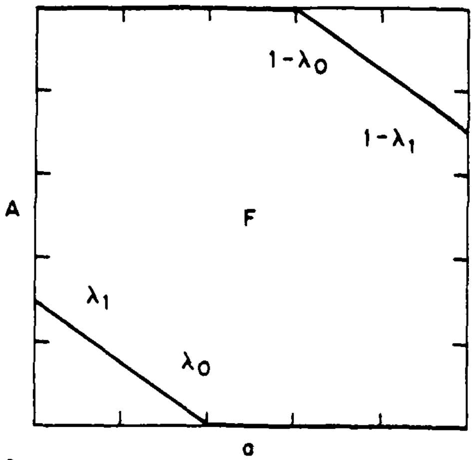
Perhaps it is more understandable to present this data in another format:

		$n=2$			$n=4$	
W	a	J	J/J_{max}	a	J	J/J_{max}
0.25	0.250	0.250	1.0	0.725	0.195	.78
0.4	0.475	0.386	.97	0.827	0.252	.63
0.5	0.618	0.455	.91	0.862	0.281	.56
0.75	0.791	0.575	.77	0.909	0.333	.44
1.0	0.866	0.652	.65	0.933	0.371	.37
1.5	0.930	0.743	.74	0.957	0.423	.42
2.0	0.957	0.797	.80	0.969	0.459	.46
2.5	0.971	0.831	.83	0.976	0.486	.49
3.0	0.979	0.856	.86	0.980	0.508	.51
4.0	0.987	0.883	.88	0.986	0.541	.54
5.0	0.992	0.909	.91	0.989	0.567	.57
6.0	0.994	0.923	.92	0.991	0.586	.59
8.0	0.997	0.941	.94	0.994	0.616	.62
10.0	0.998	0.952	.95	0.995	0.639	.64

The column labeled J/J_{max} gives the ratio of the minimum J to the quantity J_{max} defined in (18). Again, for this global minimum, $g = d = 0$.

4.2 Global Minimum

As a check that the symmetric minimum is indeed the global minimum, we have evaluated J throughout the allowed region of the six variables g, d and a, A, b, B. We could



2-88 5952A3

Fig. 3. The allowed region F for the parameter A as a function of a using the discrete three-step model.

The proof that the above minimum is indeed a global minimum follows simply by letting a and b deviate from the above values while keeping A and B as close to their optimum values as allowed by the constraint. For W small enough we have:

$$\begin{aligned} a &= b = 1-\lambda_0(1+\epsilon) \\ A &= B = 1 \end{aligned} \tag{60}$$

$$(S+T+U)_{min} \equiv -W\,(1-\lambda_0^2)\ .$$

It now follows for any positive ϵ that

$$(S+T+U)-(S+T+U)_{min} = W\,\lambda_0^2[(1+\epsilon)^2-1] \geq 0\ . \tag{61}$$

For a and b larger than their optimum values, the constraints on A and B come into play and

$$\begin{aligned} a &= b = 1-\lambda_0(1-\epsilon) \\ A &= B = 1-\lambda_1\epsilon \end{aligned} \tag{62}$$

and we find

$$\begin{aligned} &(S+T+U)-(S+T+U)_{min} = \\ &\quad 2\epsilon\lambda_1(1-\lambda_1)+(\lambda_1^2-W\lambda_0^2)\,[1-(1-a)^2] \\ &\quad \geq 0 \end{aligned} \tag{63}$$

if $(\lambda_1^2 - W\lambda_0^2)$ is positive (and if ϵ is positive, of course).

When W grows so that $(\lambda_1^2 - W\lambda_0^2)$ becomes negative, one should repeat the above procedure around the values $a = b = 1$ and $A = B = 1-\lambda_1$ to prove the global nature of the minimum in this region. Alternatively, one may argue that the feasible region is defined, in this case, by hyperplanes, so that the minimum must occur at a vertex, as given above.

7. Summary

We find that the problem of optimal design, with fusion and detector tuning, is difficult but tractable. Our simple examples yield some insight into how the best achievable cost varies between its bounds and how that best cost depends on the prior distribution and the cost function itself.

By utilising the technique of invariant imbedding, that is by considering a general class of response functions that contain the exponential response model as a particular case, we can trace a discontinuous change in design parameters, even though the optimum cost varies smoothly. We cannot yet give a detailed explanation of the critical value at which this jump occurs. The third model studied also has such a discontinuity, and permitted a continuous transition to the state of complete information (perfect discrimination). In this case, the cost depends on the degree of ambiguity in a quadratic manner.

Finally, in all cases, we found that the best cost is achieved with a symmetric choice of parameters for the individual detectors. We do not yet have a general characterization of response functions for which this is always the case regardless of costs and prior probabilities.

The problem considered here is not only of theoretical interest, but has many practical applications ranging from optimal design of complex particle detector systems to the design of seismic and warning systems.

Acknowledgements

One of the authors (PBK) acknowledges with thanks the hospitality of the Department of Operations Research, Weatherheading School of Management, Case–Western Reserve University, and its chairmen Arnold Reisman and Hamilton Emmons. We both thank Dr. Rabindar Madan of the Office of Naval Research for calling this problem to our attention.

REFERENCES

1. R. Srinivasan, *Distributed Radar Detection Theory*, IEEE Proc. F (GB) **133**, Vol. 1, pp. 55–60 (1986). See also, R. Srinivasan, Signal Processing (Netherlands) **11**, Vol. 4, pp. 319–327 (1986).
2. R. Srinivasan, P. Sharma and V. Malik, *Distributed Detection of Swerling Targets*, IEEE Proc. (GB) **133**, Vol. 7, pp. 624–629 (1986).
3. R. R. Tenney and N. R. Sandell, *Detection With Distributed Sensors*, IEEE Transactions on Aerospace and Electronic Systems **17**, Vol. 4, pp. 501–509 (1981), and **17**, Vol. 5, p. 736 (1981).
4. L. K. Ekchian and R. R. Tenney, *Recursive Solution of Distributed Detection/Communication Problems*, Proc. of the 1983 American Control Conference, Vol. 3, San Francisco, CA, pp. 1338–1339.
5. Z. Chair and P. K. Varshney, *Optimal Data Fusion in Multiple Detector Systems*, IEEE Transactions on Aerospace and Electronic Systems **22**, Vol. 1, pp. 98–101 (1986).
6. A. R. Reibman and L. W. Nolte, *Optimal Detection and Performance of Distributed Sensor Systems*, IEEE Transactions on Aerospace and Electronic Systems **23**, Vol. 1, pp. 24–30 (1987).
7. F. A. Sadjadi, *Hypothesis Testing in a Distributed Environment*, IEEE Transactions on Aerospace and Electronic Systems **22**, Vol. 2, pp. 134–137 (1986).
8. M. R. Hestenes, *Calculus of Variations and Optimal Control Theory* (New York, Wiley, 1966).
9. M. Einhorn and R. Blankenbecler, *Bounds on Scattering Amplitudes*, Annals of Physics **67**, 480 (1971).
10. The key point here is that the likelihood ratio is a monotonic function of x, so R is an interval including one endpoint.
11. M. Cherikh, *Design Discontinuities in Distributed Sensors: Implications for CFAR Systems*, in preparation.

This page is blank

APPENDIX B

REVIEW OF SOME RELATED RESULTS

Tenney and Sandell [1981] discussed team action in the case of binary hypotheses and binary actions for two stations. They selected a particular form for the cost function, established that the optimum signalling/action rule is a likelihood ratio test, and presented coupled equations determining the likelihood ratio thresholds. Numerical examples are given, showing the complexity of the problem, for several choices of the detector response functions f(y|h). The cost function is described by a single parameter, the cost of having both stations miss. Of course, as this parameter becomes large, the optimal solution is to have both stations act oppositely, no matter what the signal.

Ekchian and Tenney [1982] extend the analysis to more general topologies. They note that the number of regions into which each station divides the space of received signals depends on how many signals it may emit, and the number of such division rules (which they call thresholds, for the binary case) is equal to the total number of signal combinations that can be received from all the other stations. They propose that an extension of the dynamic programming concept can be applied to the analysis of tree-like topologies.

Chair and Varshney [1986] turn to the fusion question, and discuss the optimum fusion rule, for a binary hypothesis-binary action situation, with binary signals. Essentially they compute the updated conditional probability that either hypothesis is true, given the signals from the stations, and their known response characteristics. Although there are some notational problems with the paper, the result is correct.

Kushner and Pacut [1982] opened the discussion of the remeasurement problem, which is the simplest example of what we call the "call-back" problem. There are two stations, which can communicate with each other, in either direction. At each station, remeasurement can reduce the probability of error, but it has some cost. They formulate the problem of deciding when it is appropriate, given the signal from the other station, to remeasure before proceeding to action. The capacity of the information channel is not clearly defined, as transmission of the full posterior probability, which they assume, might be as costly as transmission of the full signal (y) received. Again, the computations are quite complex, even for the simple exponential response function.

Srinivasan [1986] provided explicit formulas for the relation among system operating characteristics and those of the distributed sensors, and noted that the optimal tuning for the sensors depends on the choice of the fusion rule. Performance characteristics are given for systems of 2 and three detectors with slow Raleigh fading (equivalent to the case of exponential distributions of signals), and it is noted that the optimal rule for 2 detectors is BOTH (called "AND"). Srinivasan, Sharma and Malik [1986] apply the methods to 2, 3 and 4 detectors for the case of Swerling targets, and provide (semi-log) plots of the optimal performance of those systems. They note that the performance comes quite close to that of a one-sensor system receiving the same amount of information. Due to some technical assumptions, the plots that

they present for 3 and four detector systems do not make this clear. [They assume that repeat pulses of the same amplitude are received at the single site.]

Sadjadi [1986] has extended the discussion to any number of stations and hypotheses, with the number of actions equal to the number of hypotheses. The ideal decision rules would fuse the data from all stations, to arrive at a grand updated probability estimate, but each station must act in ignorance of the others. Thus the space of received signals $(y_1,....,y_N)$ is sliced by hyperplanes parallel to the axes, representing the thresholds for the individual stations. Each station has as many thresholds (defined in terms of the likelihood ratios) as are needed to label the hypotheses. That is, for any signal received, one and only one of the hypotheses will have a posterior probability larger than any of the other hypotheses. The effect of the other stations is indirect, through the fact that they all know the form of the common cost function.

C.iu [1987] has discussed a related problem, where he shows that a nominator-detector scheme can lower computational costs, by screening unlikely candidates using a low cost test ["sensor"]. He assumes, however, that, for cases which are not rejects, the nominator sensor passes full information to the second detector.

Schwartz and Pelkowitz [1987] have considered the problem of minimizing the time to reach a decision, for a given fixed False Alarm Rate (FAR). This work is not directly relevant to what we have studied up until now, but will be relevant to the question of call-back strategies, discussed below.

Reibman and Nolte [1987] have done some model calculations for a three detector system, using shifted generalized Gaussian distributions to represent the sensor response to the alternative hypotheses (that is, constant signal in generalized Gaussian noise). They correctly establish that the system in which the fusion rule and the sensor tunings are jointly optimized is superior to any system in which either of those is fixed a priori. They do not solve the optimization problem directly, as we do, but use coupled equations which must hold at the optimum. They find that the optimal tuning for the cases considered is symmetric, but do not know whether that is a general rule or an accident. [We have established (see below) that it is NOT a general rule]

Additional recent works are cited in the references.

F.4 Our Own Related Work

F.4.1. Experimental Design

Our own thinking is strongly influenced by the work of the statistician Blackwell [1957] on the optimal design of experiments. The analogy is clear. An optimal experiment minimizes the chance of mistaking the hypothesis. An optimal detection system minimizes the "cost" of mislabelling the state of the world. Blackwell was able to show, by a sophisticated application of the theory of two-person zero-sum games, that it is possible for one experimental design to completely dominate another. That is, no ma..er what the cost function, and no matter what the prior probabilities, the expected cost of following the better design is lower than the expected cost of following the other.

This is an enormously important result, because, in real applications, estimates of the cost function and the prior probabilities are little more than political and social guesswork. Hence, the possibility of establishing the superiority of a design, independent of the costs and probabilities, is extremely valuable. Blackwell's result is given in terms of a necessary and sufficient condition on the convex hull of the vectors representing the joint probabilities of actions and hypotheses, when the prior probabilities are all equal.

In our own preliminary work the ideas of Blackwell have a direct interpretation. We determine the contour of possible values of (F,M), given the response functions f(y|h). This contour is exactly the envelope of the various convex hulls defined by various choices of threshold. In particular, Blackwell's theorem states that any operating point (F,M) in the interior of the allowed region, is dominated by a point on the boundary of the region. [We had obtained this result directly, using the linearity of the cost function.] Further, from the necessity part of Blackwell's theorem, it follows that none of the operating points on the boundary of the allowed region completely dominates any of the other points on that boundary. It may, further, be shown that any given choice of the cost matrix determines an operating point, or line, on the conxev hull of that boundary.

In the proposed work we will develop these relations fully, with particular attention to two situations in which it seems likely that dominance may occur:

* comparison of topologies and signal sets
* evaluation of call-back strategies.

In a call-back strategy, each station has the option of either proceeding to signal/act, or calling back to one or more of those stations from which it receives signals, to ask for a more detailed report. It seems likely that the optimal development of call-back strategies will involve the maximum entropy principle. In particular, each station will form estimates of what the others are likely to say, based on assumptions of randomness, but subject to the signals they have already transmitted. The MEP is the best known technique for implementing this concept of constrained randomness. It has been applied to the analogous problem of information retrieval in databases, by Kantor [1985] and by Kantor and Lee [1986].

F.4.2 Related work on the Maximum Entropy Principle

The Maximum Entropy Principle (MEP) is a mathematical technique [Jaynes 1957a,1957b] for making and/or facilitating decisions in the presence of probabilistic information and constraints. [Smith, ed. 1982,83,84,85,86] A priori probabilities or prior information can also be included [Johnson and Shore 1983]. Applications of the MEP approach include such abstract topics as "good null hypothesis" in statistics, computerized "expert systems" [Cheeseman 1983], the processing of seismic data [Burg 1975], the inversion of data in geologic problems [Rietsch 1977], the enhancing of blurred photographs for astronomical and law enforcement uses [Gull and Daniel 1978], and finally to the construction of the quantum mechanical density matrix from realistic (non-ideal) data [Blankenbecler and Partovi 1985].

One of us has explored the application of the MEP to the problem of retrieving "relevant documents" from a very large data base [Kantor 1984]. Every document may be described by one or more descriptors ("keywords") while constraints

take the form of "probabilities of relevance" or "expected values." For example, the system may be told that "keyword A" carries a 70% chance of relevance, and so on. The system then gathers information about the co-occurrence of various combinations of keywords, such as "A and B but not C." The MEP uniquely determines the chance of relevance for each such combination, given the constraints on co-occurrence, and the prior estimates of properties as described above.

We have established that this information is sufficient to optimize the data retrieval process when costs are measured by any reasonable combination of man and machine time, and have given a general procedure for dealing with the problem of inconsistent prior estimates [Kantor and Lee 1986, 1987a].

The most important features of the MEP approach are the following:

1. Probabilistic information and constraint are accepted at all stages.
2. The alternatives among which a choice is to be made are presented in a well-determined rank order.
3. The calculations are completed in "real time."
4. The output of this procedure, a rank-ordering, could serve as input for another level of decision making, i.e. formulating appropriate action.

These characteristics justify examination of the potential of the MEP as a decision tool in a more general context. It is clearly important for any logic or decision making system to be able to accept probabilistic information and rank-order alternatives in real time.

The relevance of the MEP to the DSS problem goes beyond these generalities. The detailed operation of a distributed system should allow for two-way messages between the various stations. The communication structure can be represented as a matrix, whose rows and columns are labeled by the various stations. The entry at the intersection of Row I and column J describes the set of signals that Station J may send to Station I. This description includes an enumeration of the signals - "1", "2", and so on, and a specification of the corresponding regions in the y-space which trigger those various signals, depending, as we have noted, on the signals input to Station J from its neighbors.

In published work, the signals flow only one way. In particular, there are no "interrogatory signals" by which station J may ask Station I to elaborate on its report. Such elaboration is clearly possible if, for example, a continuous variable (y) has been replaced by a simple discrete signal u selected from U(J,I). The decision to request an elaboration is exactly parallel to the decision to "request a document." It must be based on an estimation of the probability that the elaborated signal will improve decision making sufficiently to offset the added cost and delay. The work by Kushner and Pacut cited above is similar in spirit, but considered only remeasurement at a single station.

Thus, it seems likely that the optimizing characteristics of an MEP rank ordering will carry over to the DSS arena. We plan to study this aspect of the problem after detailed examination of a variety of optimal control models.

APPENDIX C

2. Project Goals

The anticipated payoff of both stage one and stage two of the research is:

1. a simplification in the design and comparison of distributed sensor systems
2. an improvement in the operational effectiveness of such systems
3. a major reduction in programming complexity, leading to higher reliability at lower cost.

The detailed goal is to examine the optimal design and decision strategy to be used with distributed sensors. Particular attention will be given to the notion of a dominant system design (in the sense of Blackwell), to the non-linear problems of optimal parameter (threshold) determination, and finally to the potential of the Maximum Entropy Principle as an optimizing tool.

3. Proposed Research

General observations.

Every sensor of a distributed network receives a vector of signals ($\vec{y}$) and processes them to produce a much simpler vector of signals ($\vec{u}$), whose components may be binary data or other summary data (such as velocity and position estimates). These summaries are not to be considered as only facts about the situation observed by the sensor. They are also 'generalized keywords' describing the full set of information ($\vec{y}$) available at the sensor.

Any communicating station may evaluate those keywords ($\vec{u}$), in the light of other information available to it, and decide to request a more detailed report about the original data ($\vec{y}$). In this way, a sophisticated problem of optimal information retrieval is imbedded in the problem of designing a distributed sensor system.

The planned research involves three major areas, all bearing on the central problem of optimal design of distributed sensor systems. They are outlined below (in subsections 3.1,3.2,and 3.3).

3.1 Detailed analysis of model problems.

The invariant imbedding technique is a powerful tool for determining how the properties found in model problems depend on the structure of the problem. Invariant imbedding is illustrated in the attached paper by Blankenbecler and Kantor. In this particular application we have used it to elucidate the effect of ambiguous signals on overall system performance. We are able to continously vary from a problem where the report from each detector is unambiguous, to one where they are ambiguous to any desired degree.

In our analysis so far we find that (numerically, and in some models, analytically) the optimum operating point is the same for both of the detectors in a simple fusion system. It remains to establish whether this symmetry between the detectors is universally valid and, if so, why.

It also seems intuitively clear that increasing the size of the signal set should improve system performance. We will use models of the type already discussed to see whether this intuitive relation is supported by Blackwell's theorem. This will also lay the foundations for considering the trade-offs between communication costs and system performance.

Representative problems are:

1. What is the corresponding full solution for 3 identical detectors? Is the best rule for integration of signals always majority rule?
2. In the two detector case, what improvements result when the detectors are not identical. How do the thresholds vary?

3.2 Network topology and Blackwell dominance.

Networks may be described by the matrices of possible signals, mentioned above, and the signal/action rules by which each station selects a signal or action. The most general station receives input from the world at large, receives signals from one or more other stations, emits signals to one or more other stations, and may also take action. The possibilities may be represented as follows:

$A \longrightarrow B$	means A sends a signal to B
A	indicates that A receives external information
C	means that C takes action
$A \Longrightarrow B$	means that B signals to A and A can request elaboration from B .

With this notation the problem of Tenney and Sandell is:

$$A \qquad B$$

There is no overt communication.

The problem discussed by Kantor and Blankenbecler [attached] is:

$$A \longrightarrow C \longleftarrow B$$

All other models summarized above can be similarly represented.

A typical question about topology is to compare:

System I	$A \longrightarrow C \longleftarrow B$
System II	$A \longrightarrow B \longrightarrow C$

In system II, C is masked from A by B (unless B uses an enlarged signal set.) Intuitively, therefore, the performance of System II should never be better than that of System I. Blackwell's Theorem establishes a necessary and sufficient condition on the Blackwell vectors, if this dominance prevails. Model calculations and imbedding techniques will be applied to explore this relationship.

3.3 Call-back and information-seeking structures.

The work by Kushner and Pacut, described above, assumed that each station transmits a sufficient statistic, (its own estimate of the posterior probability). In the other models, a bare minimum signal is usually assumed. The situation has some conceptual similarity to the problem of sample size in quality control. If every item is examined, one gets the best possible result, but the cost is too high; sampling is therefore used. Furthermore it is well known that fixed block sampling does not perform as well as sequential sampling, in which the decision to sample further is based (in a Bayesian analysis) on the data obtained to date.

Given this analogy, we expect to be able to show that call back systems dominate non-call-back systems with the same size signal sets, and are more efficient than systems in which all the information obtainable by call-back is transmitted whether or not it is requested.

Example systems for this problem are:

System III A $\longrightarrow$B versus

System IV A $\Longleftarrow$B .

The Maximum Entropy Principle is a powerful technique for optimal retrieval of information, and we expect that it will apply to distributed sensors. To apply it we must consider:

1. What are the effective probabilistic constraints implied by available information on the physical characteristics of the threat environment?
2. What is the computational complexity of the optimization problem, and can it be solved in real time, given state-of-the-art computational power? Does it admit a high degree of parallel processing, especially for multiple targets?

DISTRIBUTION LIST

ADDRESSEE	DODAAD CODE	NUMBER OF COPIES UNCLASSIFIED/UNLIMITED
Scientific Officer	N00014	1
Administrative Contracting Officer	S3603A	1
Director, Naval Research Laboratory, ATTN: Code2627 Washington, D.C. 20375	N00173	1
Defense Technical Information Center Bldg. 5, Cameron Station Alexandria, Virginia 22314	S47031	12

END
DATE
FILMED
DTIC
9-88